# Table of Contents

What Is Amazon EC2? ...................................................................................................................... 1
Features of Amazon EC2 ............................................................................................................ 1
How to Get Started with Amazon EC2 ...................................................................................... 1
Related Services .......................................................................................................................... 2
Accessing Amazon EC2 ............................................................................................................. 3
Pricing for Amazon EC2 .............................................................................................................. 3
PCI DSS Compliance ................................................................................................................... 4
Basic Infrastructure ..................................................................................................................... 4
Amazon Machine Images and Instances ................................................................................... 5
Regions and Availability Zones ................................................................................................. 5
Storage ....................................................................................................................................... 6
Root Device Volume .................................................................................................................. 7
Networking and Security .......................................................................................................... 8
AWS Identity and Access Management .................................................................................... 8
Differences between Windows Server and an Amazon EC2 Windows Instance .................. 9
Designing Your Applications to Run on Amazon EC2 Windows Instances ......................... 10

Setting Up ................................................................................................................................... 12
Sign Up for AWS ....................................................................................................................... 12
Create an IAM User ................................................................................................................... 12
Create a Key Pair ....................................................................................................................... 14
Create a Virtual Private Cloud (VPC) .......................................................................................... 16
Create a Security Group .......................................................................................................... 17

Getting Started ........................................................................................................................ 20
Overview ..................................................................................................................................... 20
Prerequisites .............................................................................................................................. 21
Step 1: Launch an Instance ....................................................................................................... 21
Step 2: Connect to Your Instance ............................................................................................. 22
Step 3: Clean Up Your Instance ............................................................................................... 23
Next Steps .................................................................................................................................. 24

Best Practices ........................................................................................................................... 25
Tutorials ....................................................................................................................................... 27
Tutorial: Deploy a WordPress Blog .......................................................................................... 27
Prerequisites .............................................................................................................................. 27
Installing the Microsoft Web Platform Installer ........................................................................ 28
Installing WordPress .................................................................................................................. 28
Configuring Security Keys ........................................................................................................ 29
Configuring the Site Title and Administrator ......................................................................... 30
Making Your WordPress Site Public ............................................................................................ 30
Next Steps .................................................................................................................................. 31

Tutorial: Installing a WAMP Server .......................................................................................... 31

Tutorial: Installing a WIMP Server ............................................................................................ 34
Prerequisites .............................................................................................................................. 34
Prepare Your Instance .............................................................................................................. 34
Install the IIS web server .......................................................................................................... 35
Install MySQL and PHP ............................................................................................................ 36
Test Your Server ....................................................................................................................... 36

Tutorial: Increase the Availability of Your Application ......................................................... 37
Prerequisites .............................................................................................................................. 37
Scale and Load Balance Your Application ............................................................................... 38
Test Your Load Balancer ........................................................................................................... 40

Tutorial: Set Up a Windows HPC Cluster .................................................................................. 40
Prerequisites .............................................................................................................................. 41
Step 1: Create Security Groups ................................................................................................. 41
Step 2: Set Up Your Active Directory Domain Controller ......................................................... 44
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troubleshooting Sysprep</td>
<td>117</td>
</tr>
<tr>
<td>Instances</td>
<td>119</td>
</tr>
<tr>
<td>Instance Types</td>
<td></td>
</tr>
<tr>
<td>Available Instance Types</td>
<td>120</td>
</tr>
<tr>
<td>Hardware Specifications</td>
<td>121</td>
</tr>
<tr>
<td>Nitro-based Instances</td>
<td>122</td>
</tr>
<tr>
<td>Networking and Storage Features</td>
<td>123</td>
</tr>
<tr>
<td>Instance Limits</td>
<td>125</td>
</tr>
<tr>
<td>General Purpose Instances</td>
<td>125</td>
</tr>
<tr>
<td>Compute Optimized Instances</td>
<td>161</td>
</tr>
<tr>
<td>Memory Optimized Instances</td>
<td>165</td>
</tr>
<tr>
<td>Storage Optimized Instances</td>
<td>173</td>
</tr>
<tr>
<td>Accelerated Computing Instances</td>
<td>178</td>
</tr>
<tr>
<td>Changing the Instance Type</td>
<td>188</td>
</tr>
<tr>
<td>Instance Purchasing Options</td>
<td>191</td>
</tr>
<tr>
<td>Determining the Instance Lifecycle</td>
<td>192</td>
</tr>
<tr>
<td>On-Demand Instances</td>
<td>193</td>
</tr>
<tr>
<td>Reserved Instances</td>
<td>197</td>
</tr>
<tr>
<td>Scheduled Instances</td>
<td>227</td>
</tr>
<tr>
<td>Spot Instances</td>
<td>231</td>
</tr>
<tr>
<td>Dedicated Hosts</td>
<td>294</td>
</tr>
<tr>
<td>Dedicated Instances</td>
<td>312</td>
</tr>
<tr>
<td>On-Demand Capacity Reservations</td>
<td>317</td>
</tr>
<tr>
<td>Instance Lifecycle</td>
<td>329</td>
</tr>
<tr>
<td>Instance Launch</td>
<td>330</td>
</tr>
<tr>
<td>Instance Stop and Start (Amazon EBS-Backed Instances Only)</td>
<td>330</td>
</tr>
<tr>
<td>Instance Hibernate (Amazon EBS-Backed Instances Only)</td>
<td>331</td>
</tr>
<tr>
<td>Instance Reboot</td>
<td>331</td>
</tr>
<tr>
<td>Instance Retirement</td>
<td>331</td>
</tr>
<tr>
<td>Instance Termination</td>
<td>332</td>
</tr>
<tr>
<td>Differences Between Reboot, Stop, Hibernate, and Terminate</td>
<td>332</td>
</tr>
<tr>
<td>Launch</td>
<td>333</td>
</tr>
<tr>
<td>Connect</td>
<td>379</td>
</tr>
<tr>
<td>Stop and Start</td>
<td>383</td>
</tr>
<tr>
<td>Hibernate</td>
<td>385</td>
</tr>
<tr>
<td>Reboot</td>
<td>385</td>
</tr>
<tr>
<td>Retire</td>
<td>386</td>
</tr>
<tr>
<td>Terminate</td>
<td>388</td>
</tr>
<tr>
<td>Recover</td>
<td>393</td>
</tr>
<tr>
<td>Configure Instances</td>
<td>394</td>
</tr>
<tr>
<td>EC2Launch</td>
<td>395</td>
</tr>
<tr>
<td>EC2Config Service</td>
<td>405</td>
</tr>
<tr>
<td>PV Drivers</td>
<td>431</td>
</tr>
<tr>
<td>AWS NVMe Drivers</td>
<td>446</td>
</tr>
<tr>
<td>Optimizing CPU Options</td>
<td>447</td>
</tr>
<tr>
<td>Setting the Time</td>
<td>460</td>
</tr>
<tr>
<td>Setting the Password</td>
<td>464</td>
</tr>
<tr>
<td>Adding Windows Components</td>
<td>464</td>
</tr>
<tr>
<td>Configuring a Secondary Private IPv4 Address</td>
<td>468</td>
</tr>
<tr>
<td>Running Commands at Launch</td>
<td>472</td>
</tr>
<tr>
<td>Instance Metadata and User Data</td>
<td>478</td>
</tr>
<tr>
<td>SQL Server Clustering in EC2</td>
<td>489</td>
</tr>
<tr>
<td>Upgrade Windows Instances</td>
<td>495</td>
</tr>
<tr>
<td>Performing a Server Migration</td>
<td>496</td>
</tr>
<tr>
<td>Performing an In-Place Upgrade</td>
<td>496</td>
</tr>
<tr>
<td>Performing an Automated Upgrade</td>
<td>500</td>
</tr>
<tr>
<td>Migrating to Latest Generation Instance Types</td>
<td>506</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Network and Security</td>
<td>578</td>
</tr>
<tr>
<td>Monitoring</td>
<td>532</td>
</tr>
<tr>
<td>Elastic Graphics</td>
<td>520</td>
</tr>
<tr>
<td>Security Groups</td>
<td>583</td>
</tr>
<tr>
<td>Key Pairs</td>
<td>578</td>
</tr>
<tr>
<td>Pricing for Elastic Graphics</td>
<td>522</td>
</tr>
<tr>
<td>Elastic Graphics Limitations</td>
<td>522</td>
</tr>
<tr>
<td>Working with Elastic Graphics</td>
<td>522</td>
</tr>
<tr>
<td>Security Group Rules</td>
<td>584</td>
</tr>
<tr>
<td>Connecting to Your Windows Instance if You Lose Your Private Key</td>
<td>583</td>
</tr>
<tr>
<td>Deleting Your Key Pair</td>
<td>582</td>
</tr>
<tr>
<td>Retrieving the Public Key for Your Key Pair From Your Instance</td>
<td>581</td>
</tr>
<tr>
<td>Default Security Groups</td>
<td>586</td>
</tr>
<tr>
<td>Automated and Manual Monitoring</td>
<td>533</td>
</tr>
<tr>
<td>Best Practices for Monitoring</td>
<td>534</td>
</tr>
<tr>
<td>Monitoring the Status of Your Instances</td>
<td>535</td>
</tr>
<tr>
<td>Scheduled Events</td>
<td>539</td>
</tr>
<tr>
<td>Monitoring Your Instances Using CloudWatch</td>
<td>546</td>
</tr>
<tr>
<td>Enable Detailed Monitoring</td>
<td>546</td>
</tr>
<tr>
<td>List Available Metrics</td>
<td>548</td>
</tr>
<tr>
<td>Get Statistics for Metrics</td>
<td>557</td>
</tr>
<tr>
<td>Graph Metrics</td>
<td>564</td>
</tr>
<tr>
<td>Create an Alarm</td>
<td>564</td>
</tr>
<tr>
<td>Create Alarms That Stop, Terminate, Reboot, or Recover an Instance</td>
<td>565</td>
</tr>
<tr>
<td>Automating Amazon EC2 with CloudWatch Events</td>
<td>574</td>
</tr>
<tr>
<td>Logging API Calls with AWS CloudTrail</td>
<td>574</td>
</tr>
<tr>
<td>Amazon EC2 and Amazon EBS Information in CloudTrail</td>
<td>575</td>
</tr>
<tr>
<td>Understanding Amazon EC2 and Amazon EBS Log File Entries</td>
<td>575</td>
</tr>
<tr>
<td>Auditing Users that Connect via EC2 Instance Connect</td>
<td>576</td>
</tr>
<tr>
<td>Monitoring Your Instances Using CloudWatch</td>
<td>546</td>
</tr>
<tr>
<td>Enable Detailed Monitoring</td>
<td>546</td>
</tr>
<tr>
<td>List Available Metrics</td>
<td>548</td>
</tr>
<tr>
<td>Get Statistics for Metrics</td>
<td>557</td>
</tr>
<tr>
<td>Graph Metrics</td>
<td>564</td>
</tr>
<tr>
<td>Create an Alarm</td>
<td>564</td>
</tr>
<tr>
<td>Create Alarms That Stop, Terminate, Reboot, or Recover an Instance</td>
<td>565</td>
</tr>
<tr>
<td>Automating Amazon EC2 with CloudWatch Events</td>
<td>574</td>
</tr>
<tr>
<td>Logging API Calls with AWS CloudTrail</td>
<td>574</td>
</tr>
<tr>
<td>Amazon EC2 and Amazon EBS Information in CloudTrail</td>
<td>575</td>
</tr>
<tr>
<td>Understanding Amazon EC2 and Amazon EBS Log File Entries</td>
<td>575</td>
</tr>
<tr>
<td>Auditing Users that Connect via EC2 Instance Connect</td>
<td>576</td>
</tr>
<tr>
<td>Identifying the Instance Identity Document</td>
<td>518</td>
</tr>
<tr>
<td>Inspecting the System UUID</td>
<td>519</td>
</tr>
<tr>
<td>Using CloudWatch Metrics to Monitor Elastic Graphics</td>
<td>527</td>
</tr>
<tr>
<td>Elastic Graphics Metrics</td>
<td>527</td>
</tr>
<tr>
<td>Elastic Graphics Dimensions</td>
<td>528</td>
</tr>
<tr>
<td>Creating CloudWatch Alarms to Monitor Elastic Graphics</td>
<td>528</td>
</tr>
<tr>
<td>Troubleshooting</td>
<td>529</td>
</tr>
<tr>
<td>Investigating Application Performance Issues</td>
<td>529</td>
</tr>
<tr>
<td>Resolving Unhealthy Status Issues</td>
<td>531</td>
</tr>
<tr>
<td>Monitoring Your Instances Using CloudWatch</td>
<td>546</td>
</tr>
<tr>
<td>Enable Detailed Monitoring</td>
<td>546</td>
</tr>
<tr>
<td>List Available Metrics</td>
<td>548</td>
</tr>
<tr>
<td>Get Statistics for Metrics</td>
<td>557</td>
</tr>
<tr>
<td>Graph Metrics</td>
<td>564</td>
</tr>
<tr>
<td>Create an Alarm</td>
<td>564</td>
</tr>
<tr>
<td>Create Alarms That Stop, Terminate, Reboot, or Recover an Instance</td>
<td>565</td>
</tr>
<tr>
<td>Automating Amazon EC2 with CloudWatch Events</td>
<td>574</td>
</tr>
<tr>
<td>Logging API Calls with AWS CloudTrail</td>
<td>574</td>
</tr>
<tr>
<td>Amazon EC2 and Amazon EBS Information in CloudTrail</td>
<td>575</td>
</tr>
<tr>
<td>Understanding Amazon EC2 and Amazon EBS Log File Entries</td>
<td>575</td>
</tr>
<tr>
<td>Auditing Users that Connect via EC2 Instance Connect</td>
<td>576</td>
</tr>
<tr>
<td>Troubleshooting</td>
<td>Page</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>Checking the Import Task Status</td>
<td>988</td>
</tr>
<tr>
<td>Backing Up Your Imported Instance</td>
<td>989</td>
</tr>
<tr>
<td>Troubleshooting</td>
<td>989</td>
</tr>
<tr>
<td>Error: Add-in cannot be installed</td>
<td>989</td>
</tr>
<tr>
<td>Installation Errors</td>
<td>989</td>
</tr>
<tr>
<td>Checking the Log File</td>
<td>990</td>
</tr>
<tr>
<td>Errors Importing a VM</td>
<td>990</td>
</tr>
<tr>
<td>Uninstalling the Add-In</td>
<td>990</td>
</tr>
<tr>
<td>AWS Management Pack</td>
<td>992</td>
</tr>
<tr>
<td>Overview of AWS Management Pack for System Center 2012</td>
<td>992</td>
</tr>
<tr>
<td>Overview of AWS Management Pack for System Center 2007 R2</td>
<td>994</td>
</tr>
<tr>
<td>Downloading</td>
<td>995</td>
</tr>
<tr>
<td>System Center 2012</td>
<td>995</td>
</tr>
<tr>
<td>System Center 2007 R2</td>
<td>995</td>
</tr>
<tr>
<td>Deploying</td>
<td>996</td>
</tr>
<tr>
<td>Step 1: Installing the AWS Management Pack</td>
<td>996</td>
</tr>
<tr>
<td>Step 2: Configuring the Watcher Node</td>
<td>998</td>
</tr>
<tr>
<td>Step 3: Create an AWS Run As Account</td>
<td>998</td>
</tr>
<tr>
<td>Step 4: Run the Add Monitoring Wizard</td>
<td>1001</td>
</tr>
<tr>
<td>Step 5: Configure Ports and Endpoints</td>
<td>1005</td>
</tr>
<tr>
<td>Using</td>
<td>1005</td>
</tr>
<tr>
<td>Views</td>
<td>1006</td>
</tr>
<tr>
<td>Discoveries</td>
<td>1020</td>
</tr>
<tr>
<td>Monitors</td>
<td>1021</td>
</tr>
<tr>
<td>Rules</td>
<td>1022</td>
</tr>
<tr>
<td>Events</td>
<td>1022</td>
</tr>
<tr>
<td>Health Model</td>
<td>1023</td>
</tr>
<tr>
<td>Customizing the AWS Management Pack</td>
<td>1024</td>
</tr>
<tr>
<td>Upgrading</td>
<td>1025</td>
</tr>
<tr>
<td>System Center 2012</td>
<td>1025</td>
</tr>
<tr>
<td>System Center 2007 R2</td>
<td>1026</td>
</tr>
<tr>
<td>Uninstalling</td>
<td>1026</td>
</tr>
<tr>
<td>System Center 2012</td>
<td>1026</td>
</tr>
<tr>
<td>System Center 2007 R2</td>
<td>1027</td>
</tr>
<tr>
<td>Troubleshooting</td>
<td>1027</td>
</tr>
<tr>
<td>Errors 4101 and 4105</td>
<td>1027</td>
</tr>
<tr>
<td>Error 4513</td>
<td>1027</td>
</tr>
<tr>
<td>Event 623</td>
<td>1028</td>
</tr>
<tr>
<td>Events 2023 and 2120</td>
<td>1028</td>
</tr>
<tr>
<td>Event 6024</td>
<td>1028</td>
</tr>
<tr>
<td>General Troubleshooting for System Center 2012 — Operations Manager</td>
<td>1028</td>
</tr>
<tr>
<td>General Troubleshooting for System Center 2007 R2</td>
<td>1029</td>
</tr>
<tr>
<td>EC2Rescue for Windows Server</td>
<td>1030</td>
</tr>
<tr>
<td>Using the GUI</td>
<td>1030</td>
</tr>
<tr>
<td>Video Walkthrough</td>
<td>1032</td>
</tr>
<tr>
<td>Analyzing an Offline Instance</td>
<td>1032</td>
</tr>
<tr>
<td>Collecting Data from an Active Instance</td>
<td>1033</td>
</tr>
<tr>
<td>Using the Command Line</td>
<td>1033</td>
</tr>
<tr>
<td>Collect Action</td>
<td>1034</td>
</tr>
<tr>
<td>Rescue Action</td>
<td>1036</td>
</tr>
<tr>
<td>Restore Action</td>
<td>1037</td>
</tr>
<tr>
<td>Using Systems Manager</td>
<td>1038</td>
</tr>
<tr>
<td>Examples</td>
<td>1040</td>
</tr>
<tr>
<td>Troubleshooting</td>
<td>1042</td>
</tr>
<tr>
<td>Troubleshooting Launch Issues</td>
<td>1042</td>
</tr>
<tr>
<td>Instance Limit Exceeded</td>
<td>1042</td>
</tr>
<tr>
<td>Insufficient Instance Capacity</td>
<td>1043</td>
</tr>
</tbody>
</table>
What Is Amazon EC2?

Amazon Elastic Compute Cloud (Amazon EC2) provides scalable computing capacity in the Amazon Web Services (AWS) cloud. Using Amazon EC2 eliminates your need to invest in hardware up front, so you can develop and deploy applications faster. You can use Amazon EC2 to launch as many or as few virtual servers as you need, configure security and networking, and manage storage. Amazon EC2 enables you to scale up or down to handle changes in requirements or spikes in popularity, reducing your need to forecast traffic.

For more information about cloud computing, see What is Cloud Computing?

Features of Amazon EC2

Amazon EC2 provides the following features:

- Virtual computing environments, known as instances
- Preconfigured templates for your instances, known as Amazon Machine Images (AMIs), that package the bits you need for your server (including the operating system and additional software)
- Various configurations of CPU, memory, storage, and networking capacity for your instances, known as instance types
- Secure login information for your instances using key pairs (AWS stores the public key, and you store the private key in a secure place)
- Storage volumes for temporary data that’s deleted when you stop or terminate your instance, known as instance store volumes
- Persistent storage volumes for your data using Amazon Elastic Block Store (Amazon EBS), known as Amazon EBS volumes
- Multiple physical locations for your resources, such as instances and Amazon EBS volumes, known as Regions and Availability Zones
- A firewall that enables you to specify the protocols, ports, and source IP ranges that can reach your instances using security groups
- Static IPv4 addresses for dynamic cloud computing, known as Elastic IP addresses
- Metadata, known as tags, that you can create and assign to your Amazon EC2 resources
- Virtual networks you can create that are logically isolated from the rest of the AWS cloud, and that you can optionally connect to your own network, known as virtual private clouds (VPCs)

For more information about the features of Amazon EC2, see the Amazon EC2 product page.

Amazon EC2 enables you to run any compatible Windows-based solution on our high-performance, reliable, cost-effective, cloud computing platform. For more information, see Windows Server on AWS.

For more information about running your website on AWS, see Web Hosting.

How to Get Started with Amazon EC2

First, you need to get set up to use Amazon EC2. After you are set up, you are ready to complete the Getting Started tutorial for Amazon EC2. Whenever you need more information about an Amazon EC2 feature, you can read the technical documentation.
Get Up and Running

- Setting Up with Amazon EC2 (p. 12)
- Getting Started with Amazon EC2 Windows Instances (p. 20)

Basics

- Amazon EC2 Basic Infrastructure for Windows (p. 4)
- Instance Types (p. 119)
- Tags (p. 964)

Networking and Security

- Amazon EC2 Key Pairs and Windows Instances (p. 578)
- Security Groups (p. 583)
- Elastic IP Addresses (p. 674)
- Amazon EC2 and Amazon VPC (p. 721)

Storage

- Amazon EBS (p. 782)
- Instance Store (p. 917)

Working with Windows Instances

- AWS Systems Manager Run Command in the AWS Systems Manager User Guide
- Differences between Windows Server and an Amazon EC2 Windows Instance (p. 9)
- Designing Your Applications to Run on Amazon EC2 Windows Instances (p. 10)
- Getting Started with AWS: Hosting a .NET Web App

If you have questions about whether AWS is right for you, contact AWS Sales. If you have technical questions about Amazon EC2, use the Amazon EC2 forum.

Related Services

You can provision Amazon EC2 resources, such as instances and volumes, directly using Amazon EC2. You can also provision Amazon EC2 resources using other services in AWS. For more information, see the following documentation:

- Amazon EC2 Auto Scaling User Guide
- AWS CloudFormation User Guide
- AWS Elastic Beanstalk Developer Guide
- AWS OpsWorks User Guide

To automatically distribute incoming application traffic across multiple instances, use Elastic Load Balancing. For more information, see Elastic Load Balancing User Guide.
To monitor basic statistics for your instances and Amazon EBS volumes, use Amazon CloudWatch. For more information, see the Amazon CloudWatch User Guide.

To automate actions, such as activating a Lambda function whenever a new Amazon EC2 instance starts, use Amazon CloudWatch Events. For more information, see the Amazon CloudWatch Events User Guide.

To monitor the calls made to the Amazon EC2 API for your account, including calls made by the AWS Management Console, command line tools, and other services, use AWS CloudTrail. For more information, see the AWS CloudTrail User Guide.

To get a managed relational database in the cloud, use Amazon Relational Database Service (Amazon RDS) to launch a database instance. Although you can set up a database on an EC2 instance, Amazon RDS offers the advantage of handling your database management tasks, such as patching the software, backing up, and storing the backups. For more information, see Amazon Relational Database Service Developer Guide.

To import virtual machine (VM) images from your local environment into AWS and convert them into ready-to-use AMIs or instances, use VM Import/Export. For more information, see the VM Import/Export User Guide.

## Accessing Amazon EC2

Amazon EC2 provides a web-based user interface, the Amazon EC2 console. If you’ve signed up for an AWS account, you can access the Amazon EC2 console by signing into the AWS Management Console and selecting **EC2** from the console home page.

If you prefer to use a command line interface, you have the following options:

**AWS Command Line Interface (CLI)**

Provides commands for a broad set of AWS products, and is supported on Windows, Mac, and Linux. To get started, see AWS Command Line Interface User Guide. For more information about the commands for Amazon EC2, see `ec2` in the AWS CLI Command Reference.

**AWS Tools for Windows PowerShell**

Provides commands for a broad set of AWS products for those who script in the PowerShell environment. To get started, see the AWS Tools for Windows PowerShell User Guide. For more information about the cmdlets for Amazon EC2, see the AWS Tools for PowerShell Cmdlet Reference.

Amazon EC2 provides a Query API. These requests are HTTP or HTTPS requests that use the HTTP verbs GET or POST and a Query parameter named `Action`. For more information about the API actions for Amazon EC2, see Actions in the Amazon EC2 API Reference.

If you prefer to build applications using language-specific APIs instead of submitting a request over HTTP or HTTPS, AWS provides libraries, sample code, tutorials, and other resources for software developers. These libraries provide basic functions that automate tasks such as cryptographically signing your requests, retrying requests, and handling error responses, making it is easier for you to get started. For more information, see AWS SDKs and Tools.

## Pricing for Amazon EC2

When you sign up for AWS, you can get started with Amazon EC2 for free using the AWS Free Tier.
Amazon EC2 provides the following purchasing options for instances:

**On-Demand Instances**

Pay for the instances that you use by the hour, with no long-term commitments or upfront payments.

**Reserved Instances**

Make a low, one-time, up-front payment for an instance, reserve it for a one- or three-year term, and pay a significantly lower hourly rate for these instances.

**Spot Instances**

Request unused EC2 instances, which can lower your costs significantly.

For a complete list of charges and specific prices for Amazon EC2, see Amazon EC2 Pricing.

To calculate the cost of a sample provisioned environment, see Cloud Economics Center.

To see your bill, go to the Billing and Cost Management Dashboard in the AWS Billing and Cost Management console. Your bill contains links to usage reports that provide details about your bill. To learn more about AWS account billing, see AWS Account Billing.

If you have questions concerning AWS billing, accounts, and events, contact AWS Support.

For an overview of Trusted Advisor, a service that helps you optimize the costs, security, and performance of your AWS environment, see AWS Trusted Advisor.

**PCI DSS Compliance**

Amazon EC2 supports the processing, storage, and transmission of credit card data by a merchant or service provider, and has been validated as being compliant with Payment Card Industry (PCI) Data Security Standard (DSS). For more information about PCI DSS, including how to request a copy of the AWS PCI Compliance Package, see PCI DSS Level 1.

**Amazon EC2 Basic Infrastructure for Windows**

As you get started with Amazon EC2, you'll benefit from understanding the components of its basic infrastructure and how they compare or contrast with your own data centers.

**Concepts**

- Amazon Machine Images and Instances (p. 5)
- Regions and Availability Zones (p. 5)
- Storage (p. 6)
- Root Device Volume (p. 7)
- Networking and Security (p. 8)
- AWS Identity and Access Management (p. 8)
- Differences between Windows Server and an Amazon EC2 Windows Instance (p. 9)
- Designing Your Applications to Run on Amazon EC2 Windows Instances (p. 10)
Amazon Machine Images and Instances

An Amazon Machine Image (AMI) is a template that contains a software configuration (for example, an operating system, an application server, and applications). From an AMI, you launch instances, which are copies of the AMI running as virtual servers in the cloud.

Amazon publishes many AMIs that contain common software configurations for public use. In addition, members of the AWS developer community have published their own custom AMIs. You can also create your own custom AMI or AMIs; doing so enables you to quickly and easily start new instances that have everything you need. For example, if your application is a website or web service, your AMI could include a web server, the associated static content, and the code for the dynamic pages. As a result, after you launch an instance from this AMI, your web server starts, and your application is ready to accept requests.

You can launch different types of instances from a single AMI. An instance type essentially determines the hardware of the host computer used for your instance. Each instance type offers different compute and memory facilities. Select an instance type based on the amount of memory and computing power that you need for the applications or software that you plan to run on the instance. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types. You can also launch multiple instances from an AMI, as shown in the following figure.

Your Windows instances keep running until you stop or terminate them, or until they fail. If an instance fails, you can launch a new one from the AMI.

Your AWS account has a limit on the number of instances that you can have running. For more information about this limit, and how to request an increase, see How many instances can I run in Amazon EC2 in the Amazon EC2 General FAQ.

Bare Metal Instances

Amazon now offers EC2 bare metal instances which are available with the EC2 Storage Optimized Amazon EC2 I3 instance family in the form of i3.metal instances. Bare metal instances allow you to run applications that benefit from deep performance analysis tools, specialized workloads that require direct access to bare metal infrastructure, legacy workloads not supported in virtual environments, licensing-restricted Tier 1 business critical applications, as well as virtualization-secured containers.

Regions and Availability Zones

Amazon has data centers in different areas of the world (for example, North America, Europe, and Asia). Correspondingly, Amazon EC2 is available to use in different Regions. By launching instances in separate Regions, you can design your application to be closer to specific customers or to meet legal or other
requirements. Prices for Amazon EC2 usage vary by Region (for more information about pricing by Region, see Amazon EC2 Pricing).

Each Region contains multiple distinct locations called Availability Zones. Each Availability Zone is engineered to be isolated from failures in other Availability Zones, and to provide inexpensive, low-latency network connectivity to other zones in the same Region. By launching instances in separate Availability Zones, you can protect your applications from the failure of a single location.

For more information about the available Regions and Availability Zones, see Using Regions and Availability Zones in the Amazon EC2 User Guide for Linux Instances.

Storage

When using Amazon EC2, you may have data that you need to store. Amazon EC2 offers the following storage options:

- Amazon Elastic Block Store (Amazon EBS)
- Amazon EC2 Instance Store (p. 917)
- Amazon Simple Storage Service (Amazon S3)

The following figure shows the relationship between these types of storage.

Amazon EBS Volumes

Amazon EBS volumes are the recommended storage option for the majority of use cases. Amazon EBS provides your instances with persistent, block-level storage. Amazon EBS volumes are essentially hard disks that you can attach to a running instance.

Amazon EBS is especially suited for applications that require a database, a file system, or access to raw block-level storage.
As illustrated in the previous figure, you can attach multiple volumes to an instance. Also, to keep a backup copy of your data, you can create a snapshot of an EBS volume, which is stored in Amazon S3. You can create a new Amazon EBS volume from a snapshot, and attach it to another instance. You can also detach a volume from an instance and attach it to a different instance. The following figure illustrates the life cycle of an EBS volume.

For more information about Amazon EBS volumes, see Amazon Elastic Block Store (p. 782).

**Instance Store**

*Instance store* provides your instances with temporary, block-level storage. This is storage that is physically attached to the host computer. The data on an instance store volume doesn't persist when the associated instance is stopped or terminated. For a list of instance store volumes available on each supported instance type, see Instance Store Volumes (p. 918).

Instance store is an option for inexpensive temporary storage. You can use instance store volumes if you don't require data persistence. For more information about instance store volumes, see Amazon EC2 Instance Store (p. 917).

**Amazon S3**

Amazon S3 is storage for the Internet. It provides a simple web service interface that enables you to store and retrieve any amount of data from anywhere on the web. For more information about Amazon S3, see the Amazon S3 product page.

**Root Device Volume**

When you launch an instance, the *root device volume* contains the image used to boot the instance. When you launch a Windows instance, a root EBS volume is created from an EBS snapshot and attached to the instance.

By default, the root volume is deleted when the instance terminates (the `DeleteOnTermination` attribute is true). Using the console, you can change `DeleteOnTermination` when you launch an instance. To change this attribute for a running instance, you must use the command line.

**To change the root device volume of an instance to persist at launch using the console**

1. Open the Amazon EC2 console.
2. From the Amazon EC2 console dashboard, click **Launch Instance**.
3. On the **Choose an Amazon Machine Image (AMI)** page, choose the AMI to use and click **Select**.
4. Follow the wizard to complete the **Choose an Instance Type** and **Configure Instance Details** pages.
5. On the **Add Storage** page, deselect the **Delete On Termination** check box for the root volume.
6. Complete the remaining wizard pages, and then click **Launch**.

You can verify the setting by viewing details for the root device volume on the instance's details pane. Next to **Block devices**, choose the entry for the root device volume. By default, **Delete on termination** is **True**. If you change the default behavior, **Delete on termination** is **False**.

**To change the root device volume of an instance to persist using the command line**

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).
- `modify-instance-attribute` (AWS CLI)
- `Edit-EC2InstanceAttribute` (AWS Tools for Windows PowerShell)

**Networking and Security**

By default, an instance is assigned public IPv4 address only if it's launched into a default VPC. An instance that's launched into a nondefault VPC must be specifically assigned a public IPv4 address at launch, or you must modify your subnet's default public IPv4 addressing behavior.

Instances can fail or terminate for reasons outside of your control. If one fails and you launch a replacement instance, the replacement has a different public IPv4 address than the original. However, if your application needs a static IPv4 address, Amazon EC2 offers **Elastic IP addresses**. For more information, see **Amazon EC2 Instance IP Addressing** (p. 656).

You can use **security groups** to control who can access your instances. These are analogous to an inbound network firewall that enables you to specify the protocols, ports, and source IP ranges that are allowed to reach your instances. You can create multiple security groups and assign different rules to each group. You can then assign each instance to one or more security groups, and we use the rules to determine which traffic is allowed to reach the instance. You can configure a security group so that only specific IP addresses or specific security groups have access to the instance. For more information, see **Amazon EC2 Security Groups for Windows Instances** (p. 583).

**AWS Identity and Access Management**

AWS Identity and Access Management (IAM) enables you to do the following:
- Create users and groups under your AWS account
- Assign unique security credentials to each user under your AWS account
- Control each user's permissions to perform tasks using AWS resources
- Allow the users in another AWS account to share your AWS resources
- Create roles for your AWS account and define the users or services that can assume them
- Use existing identities for your enterprise to grant permissions to perform tasks using AWS resources

By using IAM with Amazon EC2, you can control whether users in your organization can perform a task using specific Amazon EC2 API actions and whether they can use specific AWS resources.
Differences between Windows Server and an Amazon EC2 Windows Instance

After you launch your Amazon EC2 Windows instance, it behaves like a traditional server running Windows Server. For example, both Windows Server and an Amazon EC2 instance can be used to run your web applications, conduct batch processing, or manage applications requiring large-scale computations. However, there are important differences between the server hardware model and the cloud computing model. The way an Amazon EC2 instance runs is not the same as the way a traditional server running Windows Server runs.

Before you begin launching Amazon EC2 Windows instances, you should be aware that the architecture of applications running on cloud servers can differ significantly from the architecture for traditional application models running on your hardware. Implementing applications on cloud servers requires a shift in your design process.

The following table describes some key differences between Windows Server and an Amazon EC2 Windows instance.

<table>
<thead>
<tr>
<th>Windows Server</th>
<th>Amazon EC2 Windows Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resources and capacity are physically limited.</td>
<td>Resources and capacity are scalable.</td>
</tr>
<tr>
<td>You pay for the infrastructure, even if you don't</td>
<td>You pay for the usage of the infrastructure. We stop charging you for the</td>
</tr>
<tr>
<td>use it.</td>
<td>instance as soon as you stop or terminate it.</td>
</tr>
<tr>
<td>Occupies physical space and must be maintained on a</td>
<td>Doesn't occupy physical space and does not require regular maintenance.</td>
</tr>
<tr>
<td>regular basis.</td>
<td></td>
</tr>
<tr>
<td>Starts with push of the power button (known as</td>
<td>Starts with the launch of the instance.</td>
</tr>
<tr>
<td>cold booting).</td>
<td></td>
</tr>
<tr>
<td>You can keep the server running until it is time to</td>
<td>You can keep the server running, or stop and restart it (during which the instance</td>
</tr>
<tr>
<td>shut it down, or put it in a sleep or hibernation</td>
<td>is moved to a new host computer).</td>
</tr>
<tr>
<td>state (during which the server is powered down).</td>
<td></td>
</tr>
<tr>
<td>When you shut down the server, all resources</td>
<td>When you terminate the instance, its infrastructure is no longer available to you.</td>
</tr>
<tr>
<td>remain intact and in the state they were in when</td>
<td>You can't connect to or restart an instance after you've terminated it. However,</td>
</tr>
<tr>
<td>you switched it off. Information you stored on the</td>
<td>you can create an image from your instance while it's running, and launch new</td>
</tr>
<tr>
<td>hard drives persists and can be accessed whenever</td>
<td>instances from the image at any time.</td>
</tr>
<tr>
<td>it's needed. You can restore the server to the</td>
<td></td>
</tr>
<tr>
<td>running state by powering it on.</td>
<td></td>
</tr>
</tbody>
</table>

A traditional server running Windows Server goes through the states shown in the following diagram.
An Amazon EC2 Windows instance is similar to the traditional Windows Server, as you can see by comparing the following diagram with the previous diagram for Windows Server. After you launch an instance, it briefly goes into the pending state while registration takes place, then it goes into the running state. The instance remains active until you stop or terminate it. You can't restart an instance after you terminate it. You can create a backup image of your instance while it's running, and launch a new instance from that backup image.

Designing Your Applications to Run on Amazon EC2 Windows Instances

It is important that you consider the differences mentioned in the previous section when you design your applications to run on Amazon EC2 Windows instances.

Applications built for Amazon EC2 use the underlying computing infrastructure on an as-needed basis. They draw on necessary resources (such as storage and computing) on demand in order to perform a job, and relinquish the resources when done. In addition, they often dispose of themselves after the job is done. While in operation, the application scales up and down elastically based on resource requirements. An application running on an Amazon EC2 instance can terminate and recreate the various components at will in case of infrastructure failures.

When designing your Windows applications to run on Amazon EC2, you can plan for rapid deployment and rapid reduction of compute and storage resources, based on your changing needs.

When you run an Amazon EC2 Windows instance, you don't need to provision the exact system package of hardware, software, and storage, the way you do with Windows Server. Instead, you can focus on using a variety of cloud resources to improve the scalability and overall performance of your Windows application.

With Amazon EC2, designing for failure and outages is an integral and crucial part of the architecture. As with any scalable and redundant system, architecture of your system should account for computing, network, and storage failures. You have to build mechanisms in your applications that can handle...
different kinds of failures. The key is to build a modular system with individual components that are not tightly coupled, can interact asynchronously, and treat one another as black boxes that are independently scalable. Thus, if one of your components fails or is busy, you can launch more instances of that component without breaking your current system.

Another key element to designing for failure is to distribute your application geographically. Replicating your application across geographically distributed Regions improves high availability in your system.

Amazon EC2 infrastructure is programmable and you can use scripts to automate the deployment process, to install and configure software and applications, and to bootstrap your virtual servers.

You should implement security in every layer of your application architecture running on an Amazon EC2 Windows instance. If you are concerned about storing sensitive and confidential data within your Amazon EC2 environment, you should encrypt the data before uploading it.
Setting Up with Amazon EC2

If you've already signed up for Amazon Web Services (AWS), you can start using Amazon EC2 immediately. You can open the Amazon EC2 console, choose Launch Instance, and follow the steps in the launch wizard to launch your first instance.

If you haven't signed up for AWS yet, or if you need assistance launching your first instance, complete the following tasks to get set up to use Amazon EC2:

1. Sign Up for AWS (p. 12)
2. Create an IAM User (p. 12)
3. Create a Key Pair (p. 14)
4. Create a Virtual Private Cloud (VPC) (p. 16)
5. Create a Security Group (p. 17)

Sign Up for AWS

When you sign up for Amazon Web Services (AWS), your AWS account is automatically signed up for all services in AWS, including Amazon EC2. You are charged only for the services that you use.

With Amazon EC2, you pay only for what you use. If you are a new AWS customer, you can get started with Amazon EC2 for free. For more information, see AWS Free Tier.

If you have an AWS account already, skip to the next task. If you don't have an AWS account, use the following procedure to create one.

To create an AWS account

2. Follow the online instructions.

Part of the sign-up procedure involves receiving a phone call and entering a verification code on the phone keypad.

Note your AWS account number, because you'll need it for the next task.

Create an IAM User

Services in AWS, such as Amazon EC2, require that you provide credentials when you access them, so that the service can determine whether you have permission to access its resources. The console requires your password. You can create access keys for your AWS account to access the command line interface or API. However, we don't recommend that you access AWS using the credentials for your AWS account; we recommend that you use AWS Identity and Access Management (IAM) instead. Create an IAM user, and then add the user to an IAM group with administrative permissions or grant this user administrative permissions. You can then access AWS using a special URL and the credentials for the IAM user.
If you signed up for AWS but have not created an IAM user for yourself, you can create one using the IAM console. If you aren't familiar with using the console, see Working with the AWS Management Console for an overview.

To create an administrator user for yourself and add the user to an administrators group (console)

1. Use your AWS account email address and password to sign in as the AWS account root user to the IAM console at https://console.aws.amazon.com/iam/.

   Note
   We strongly recommend that you adhere to the best practice of using the Administrator IAM user below and securely lock away the root user credentials. Sign in as the root user only to perform a few account and service management tasks.

2. In the navigation pane, choose Users and then choose Add user.

3. For User name, enter Administrator.

4. Select the check box next to AWS Management Console access. Then select Custom password, and then enter your new password in the text box.

5. (Optional) By default, AWS requires the new user to create a new password when first signing in. You can clear the check box next to User must create a new password at next sign-in to allow the new user to reset their password after they sign in.

6. Choose Next: Permissions.

7. Under Set permissions, choose Add user to group.

8. Choose Create group.

9. In the Create group dialog box, for Group name enter Administrators.

10. Choose Filter policies, and then select AWS managed -job function to filter the table contents.

11. In the policy list, select the check box for AdministratorAccess. Then choose Create group.

   Note
   You must activate IAM user and role access to Billing before you can use the AdministratorAccess permissions to access the AWS Billing and Cost Management console. To do this, follow the instructions in step 1 of the tutorial about delegating access to the billing console.

12. Back in the list of groups, select the check box for your new group. Choose Refresh if necessary to see the group in the list.

13. Choose Next: Tags.

14. (Optional) Add metadata to the user by attaching tags as key-value pairs. For more information about using tags in IAM, see Tagging IAM Entities in the IAM User Guide.

15. Choose Next: Review to see the list of group memberships to be added to the new user. When you are ready to proceed, choose Create user.

You can use this same process to create more groups and users and to give your users access to your AWS account resources. To learn about using policies that restrict user permissions to specific AWS resources, see Access Management and Example Policies.

To sign in as this new IAM user, sign out of the AWS console, then use the following URL, where your_aws_account_id is your AWS account number without the hyphens (for example, if your AWS account number is 1234–5678–9012, your AWS account ID is 123456789012):

https://your_aws_account_id.signin.aws.amazon.com/console/

Enter the IAM user name (not your email address) and password that you just created. When you're signed in, the navigation bar displays "your_user_name @ your_aws_account_id".
Create a Key Pair

AWS uses public-key cryptography to secure the login information for your instance. You specify the name of the key pair when you launch your instance, then provide the private key to obtain the administrator password for your Windows instance so you can log in using RDP.

If you haven't created a key pair already, you can create one using the Amazon EC2 console. Note that if you plan to launch instances in multiple regions, you'll need to create a key pair in each region. For more information about regions, see Regions and Availability Zones (p. 5).

To create a key pair

1. Sign in to AWS using the URL that you created in the previous section.
2. From the AWS dashboard, choose EC2 to open the Amazon EC2 console.
3. From the navigation bar, select a region for the key pair. You can select any region that's available to you, regardless of your location. However, key pairs are specific to a region; for example, if you plan to launch an instance in the US East (Ohio) Region, you must create a key pair for the instance in the US East (Ohio) Region.
4. In the navigation pane, under **NETWORK & SECURITY**, choose **Key Pairs**.

**Tip**
The navigation pane is on the left side of the console. If you do not see the pane, it might be minimized; choose the arrow to expand the pane. You may have to scroll down to see the **Key Pairs** link.
5. Choose **Create Key Pair**.

6. Enter a name for the new key pair in the **Key pair name** field of the **Create Key Pair** dialog box, and then choose **Create**. Use a name that is easy for you to remember, such as your IAM user name, followed by `-key-pair`, plus the region name. For example, *me-key-pair-useast2*.

7. The private key file is automatically downloaded by your browser. The base file name is the name you specified as the name of your key pair, and the file name extension is `.pem`. Save the private key file in a safe place.

   **Important**
   
   This is the only chance for you to save the private key file. You'll need to provide the name of your key pair when you launch an instance and the corresponding private key each time you connect to the instance.

For more information, see Amazon EC2 Key Pairs and Windows Instances (p. 578).

---

### Create a Virtual Private Cloud (VPC)

Amazon VPC enables you to launch AWS resources into a virtual network that you’ve defined, known as a *virtual private cloud* (VPC). The newer EC2 instance types require that you launch your instances in a VPC. If you have a default VPC, you can skip this section and move to the next task, Create a Security Group (p. 17). To determine whether you have a default VPC, open the Amazon EC2 console and look for **Default VPC** under **Account Attributes** on the dashboard. If you do not have a default VPC listed on the dashboard, you can create a nondefault VPC using the steps below.

**To create a nondefault VPC**

1. Open the Amazon VPC console at [https://console.aws.amazon.com/vpc/](https://console.aws.amazon.com/vpc/).
2. From the navigation bar, select a region for the VPC. VPCs are specific to a region, so you should select the same region in which you created your key pair.
3. On the VPC dashboard, choose **Launch VPC Wizard**.
4. On the **Step 1: Select a VPC Configuration** page, ensure that **VPC with a Single Public Subnet** is selected, and choose **Select**.
5. On the **Step 2: VPC with a Single Public Subnet** page, enter a friendly name for your VPC in the **VPC name** field. Leave the other default configuration settings, and choose **Create VPC**. On the confirmation page, choose **OK**.

For more information about VPCs, see the Amazon VPC User Guide.
Create a Security Group

Security groups act as a firewall for associated instances, controlling both inbound and outbound traffic at the instance level. You must add rules to a security group that enable you to connect to your instance from your IP address using RDP. You can also add rules that allow inbound and outbound HTTP and HTTPS access from anywhere.

Note that if you plan to launch instances in multiple regions, you'll need to create a security group in each region. For more information about regions, see Regions and Availability Zones (p. 5).

Prerequisites

You'll need the public IPv4 address of your local computer. The security group editor in the Amazon EC2 console can automatically detect the public IPv4 address for you. Alternatively, you can use the search phrase "what is my IP address" in an Internet browser, or use the following service: Check IP. If you are connecting through an Internet service provider (ISP) or from behind a firewall without a static IP address, you need to find out the range of IP addresses used by client computers.

To create a security group with least privilege

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.

   Tip
   Alternatively, you can use the Amazon VPC console to create a security group. However, the instructions in this procedure don't match the Amazon VPC console. Therefore, if you switched to the Amazon VPC console in the previous section, either switch back to the Amazon EC2 console and use these instructions, or use the instructions in Set Up a Security Group for Your VPC in the Amazon VPC Getting Started Guide.

2. From the navigation bar, select a region for the security group. Security groups are specific to a region, so you should select the same region in which you created your key pair.


5. Enter a name for the new security group and a description. Use a name that is easy for you to remember, such as your IAM user name, followed by _SG_, plus the region name. For example, me_SG_uswest2.

6. In the VPC list, select your VPC. If you have a default VPC, it’s the one that is marked with an asterisk (*).

7. On the Inbound tab, create the following rules (choose Add Rule for each new rule), and then choose Create:
   - Choose HTTP from the Type list, and make sure that Source is set to Anywhere (0.0.0.0/0).
• Choose HTTPS from the Type list, and make sure that Source is set to Anywhere (0.0.0.0/0).
• Choose RDP from the Type list. In the Source box, choose My IP to automatically populate the field with the public IPv4 address of your local computer. Alternatively, choose Custom and specify the public IPv4 address of your computer or network in CIDR notation. To specify an individual IP address in CIDR notation, add the routing suffix /32, for example, 203.0.113.25/32. If your company allocates addresses from a range, specify the entire range, such as 203.0.113.0/24.

Warning
For security reasons, we don't recommend that you allow RDP access from all IPv4 addresses (0.0.0.0/0) to your instance, except for testing purposes and only for a short time.

For more information, see Amazon EC2 Security Groups for Windows Instances (p. 583).
Getting Started with Amazon EC2 Windows Instances

Let's get started with Amazon Elastic Compute Cloud (Amazon EC2) by launching, connecting to, and using a Windows instance. An instance is a virtual server in the AWS cloud. With Amazon EC2, you can set up and configure the operating system and applications that run on your instance.

When you sign up for AWS, you can get started with Amazon EC2 using the AWS Free Tier. If you created your AWS account less than 12 months ago, and have not already exceeded the free tier benefits for Amazon EC2, it will not cost you anything to complete this tutorial, because we help you select options that are within the free tier benefits. Otherwise, you'll incur the standard Amazon EC2 usage fees from the time that you launch the instance until you terminate the instance (which is the final task of this tutorial), even if it remains idle.

Contents

- Overview (p. 20)
- Prerequisites (p. 21)
- Step 1: Launch an Instance (p. 21)
- Step 2: Connect to Your Instance (p. 22)
- Step 3: Clean Up Your Instance (p. 23)
- Next Steps (p. 24)

Overview

The instance is an Amazon EBS-backed instance (meaning that the root volume is an EBS volume). You can either specify the Availability Zone in which your instance runs, or let Amazon EC2 select an Availability Zone for you. When you launch your instance, you secure it by specifying a key pair and security group. When you connect to your instance, you must specify the private key of the key pair that you specified when launching your instance.

Tasks

To complete this tutorial, perform the following tasks:

1. Launch an Instance (p. 21)
2. Connect to Your Instance (p. 22)
Prerequisites

Before you begin, be sure that you've completed the steps in Setting Up with Amazon EC2 (p. 12).

Step 1: Launch an Instance

You can launch a Windows instance using the AWS Management Console as described in the following procedure. This tutorial is intended to help you launch your first instance quickly, so it doesn't cover all possible options. For more information about the advanced options, see Launching an Instance.

To launch an instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the console dashboard, choose Launch Instance.
3. The Choose an Amazon Machine Image (AMI) page displays a list of basic configurations, called Amazon Machine Images (AMIs), that serve as templates for your instance. Select the AMI for Windows Server 2016 Base or later. Notice that these AMIs are marked "Free tier eligible."
4. On the Choose an Instance Type page, you can select the hardware configuration of your instance. Select the t2.micro type, which is selected by default. Notice that this instance type is eligible for the free tier.
5. Choose Review and Launch to let the wizard complete the other configuration settings for you.
6. On the Review Instance Launch page, under Security Groups, you'll see that the wizard created and selected a security group for you. You can use this security group, or alternatively you can select the security group that you created when getting set up using the following steps:

   a. Choose Edit security groups.
   b. On the Configure Security Group page, ensure that Select an existing security group is selected.
   c. Select your security group from the list of existing security groups, and then choose Review and Launch.
8. When prompted for a key pair, select Choose an existing key pair, then select the key pair that you created when setting up.

Alternatively, you can create a new key pair. Select Create a new key pair, enter a name for the key pair, and then choose Download Key Pair. This is the only chance for you to save the private key file, so be sure to download it. Save the private key file in a safe place. You'll need to provide the name of your key pair when you launch an instance and the corresponding private key each time you connect to the instance.

Warning
Don't select the Proceed without a key pair option. If you launch your instance without a key pair, then you can't connect to it.
When you are ready, select the acknowledgement check box, and then choose Launch Instances.

9. A confirmation page lets you know that your instance is launching. Choose View Instances to close the confirmation page and return to the console.

10. On the Instances screen, you can view the status of the launch. It takes a short time for an instance to launch. When you launch an instance, its initial state is pending. After the instance starts, its state changes to running and it receives a public DNS name. (If the Public DNS (IPv4) column is hidden, choose Show/Hide Columns (the gear-shaped icon) in the top right corner of the page and then select Public DNS (IPv4).)

11. It can take a few minutes for the instance to be ready so that you can connect to it. Check that your instance has passed its status checks; you can view this information in the Status Checks column.

Step 2: Connect to Your Instance

To connect to a Windows instance, you must retrieve the initial administrator password (see step 2 below) and then specify this password when you connect to your instance using Remote Desktop.

The name of the administrator account depends on the language of the operating system. For example, for English, it's Administrator, for French it's Administrateur, and for Portuguese it's Administrador. For more information, see Localized Names for Administrator Account in Windows in the Microsoft TechNet Wiki.

If you've joined your instance to a domain, you can connect to your instance using domain credentials you've defined in AWS Directory Service. On the Remote Desktop login screen, instead of using the local computer name and the generated password, use the fully-qualified user name for the administrator (for example, corp.example.com\Admin) and the password for this account.

The license for the Windows Server operating system (OS) allows two simultaneous remote connections for administrative purposes. The license for Windows Server is included in the price of your Windows instance. If you need more than two simultaneous remote connections, you must purchase a Remote Desktop Services (RDS) license. If you attempt a third connection, an error occurs. For more information, see Configure the Number of Simultaneous Remote Connections Allowed for a Connection.

To connect to your Windows instance using an RDP client

1. In the Amazon EC2 console, select the instance, and then choose Connect.
2. In the Connect To Your Instance dialog box, choose Get Password (it will take a few minutes after the instance is launched before the password is available).
3. Choose Browse and navigate to the private key file you created when you launched the instance. Select the file and choose Open to copy the entire contents of the file into the Contents field.
4. Choose Decrypt Password. The console displays the default administrator password for the instance in the Connect To Your Instance dialog box, replacing the link to Get Password shown previously with the actual password.
5. Record the default administrator password, or copy it to the clipboard. You need this password to connect to the instance.
6. Choose Download Remote Desktop File. Your browser prompts you to either open or save the .rdp file. Either option is fine. When you have finished, you can choose Close to dismiss the Connect To Your Instance dialog box.
   - If you opened the .rdp file, you'll see the Remote Desktop Connection dialog box.
   - If you saved the .rdp file, navigate to your downloads directory, and open the .rdp file to display the dialog box.
7. You may get a warning that the publisher of the remote connection is unknown. You can continue to connect to your instance.
8. When prompted, log in to the instance, using the administrator account for the operating system and the password that you recorded or copied previously. If your Remote Desktop Connection already has an administrator account set up, you might have to choose the Use another account option and type the user name and password manually.

   **Note**
   Sometimes copying and pasting content can corrupt data. If you encounter a "Password Failed" error when you log in, try typing in the password manually.

9. Due to the nature of self-signed certificates, you may get a warning that the security certificate could not be authenticated. Use the following steps to verify the identity of the remote computer, or simply choose Yes or Continue to continue if you trust the certificate.

   a. If you are using Remote Desktop Connection from a Windows PC, choose View certificate. If you are using Microsoft Remote Desktop on a Mac, choose Show Certificate.

   b. Choose the Details tab, and scroll down to the Thumbprint entry on a Windows PC, or the SHA1 Fingerprints entry on a Mac. This is the unique identifier for the remote computer’s security certificate.

   c. In the Amazon EC2 console, select the instance, choose Actions, and then choose Get System Log.

   d. In the system log output, look for an entry labeled RDPCERTIFICATE-THUMBPRINT. If this value matches the thumbprint or fingerprint of the certificate, you have verified the identity of the remote computer.

   e. If you are using Remote Desktop Connection from a Windows PC, return to the Certificate dialog box and choose OK. If you are using Microsoft Remote Desktop on a Mac, return to the Verify Certificate and choose Continue.

   f. [Windows] Choose Yes in the Remote Desktop Connection window to connect to your instance. [Mac OS] Log in as prompted, using the default administrator account and the default administrator password that you recorded or copied previously. Note that you might need to switch spaces to see the login screen. For more information about spaces, see support.apple.com/en-us/HT204100.

   g. If you receive an error while attempting to connect to your instance, see Remote Desktop Can’t Connect to the Remote Computer (p. 1045).

---

**Step 3: Clean Up Your Instance**

After you’ve finished with the instance that you created for this tutorial, you should clean up by terminating the instance. If you want to do more with this instance before you clean up, see Next Steps (p. 24).

**Important**
Terminating an instance effectively deletes it; you can’t reconnect to an instance after you’ve terminated it.

If you launched an instance that is not within the AWS Free Tier, you’ll stop incurring charges for that instance as soon as the instance status changes to shutting down or terminated. If you’d like to keep your instance for later, but not incur charges, you can stop the instance now and then start it again later. For more information, see Stopping Instances.

**To terminate your instance**

1. In the navigation pane, choose Instances. In the list of instances, select the instance.
2. Choose Actions, Instance State, Terminate.
3. Choose Yes, Terminate when prompted for confirmation.
Next Steps

After you start your instance, you might want to try some of the following exercises:

- Learn how to remotely manage your EC2 instance using Run Command. For more information, see AWS Systems Manager Run Command in the AWS Systems Manager User Guide.
- Configure a CloudWatch alarm to notify you if your usage exceeds the Free Tier. For more information, see Create a Billing Alarm in the AWS Billing and Cost Management User Guide.
- Add an EBS volume. For more information, see Creating an Amazon EBS Volume (p. 800) and Attaching an Amazon EBS Volume to an Instance (p. 803).
- Install the WAMP or WIMP stack. For more information, see Tutorial: Installing a WAMP Server on an Amazon EC2 Instance Running Windows Server (p. 31) and Tutorial: Installing a WIMP Server on an Amazon EC2 Instance Running Windows Server (p. 34).
Best Practices for Windows on Amazon EC2

This list of practices will help you get the best results from running Windows on Amazon EC2.

Update Windows Drivers

Maintain the latest drivers on all Windows EC2 instances to ensure the latest issue fixes and performance enhancements are applied across your fleet. Depending on your instance type, you should update AWS PV, ENA, and NVMe drivers.

- Leverage Trusted Advisor to keep Amazon EC2 Windows up to date with AWS-provided Windows drivers.
- Use SNS topics to receive updates for new driver releases.
- Use the AWS Systems Manager SSM document AWSSupport-UpgradeWindowsAWSDrivers to easily apply the updates across your instances.

Launch New Instances with the Latest Windows AMIs

AWS releases new Windows AMIs each month, which contain the latest OS patches, drivers, and launch agents. You should leverage the latest AMI when you launch new instances or when you build your own custom images.

- To build with the latest available AMIs, see Query for the Latest Windows AMI Using Systems Manager Parameter Store.

Test System/Application Performance Before Migration

Migrating enterprise applications to AWS can involve many variables and configurations. Always performance test the EC2 solution to ensure that:

- Instance types are properly configured, including instance size, enhanced networking, and tenancy (shared or dedicated).
- Instance topology is appropriate for the workload and leverages high-performance features when necessary (dedicated tenancy, placement groups, instance store volumes, bare metal).

Update Launch Agents

Update to the latest EC2Config (2012 R2 and earlier) or EC2Launch (2016 and later) agents to ensure the latest issue fixes are applied across your fleet.

- For EC2Config update instructions, see Installing the Latest Version of EC2Config. Leverage Trusted Advisor to describe all instances running older versions of EC2Config. You can use SNS topics to receive updates for new releases.
- For EC2Launch update instructions, see Installing the Latest Version of EC2Launch.

Security and Network

- Manage access to AWS resources and APIs using identity federation, IAM users, and IAM roles. Establish credential management policies and procedures for creating, distributing, rotating, and revoking AWS access credentials. For more information, see IAM Best Practices in the IAM User Guide.
• Implement the least permissive rules for your security group. For more information, see Security Group Rules (p. 584).

• Regularly patch, update, and secure the operating system and applications on your instance. For more information about updating Amazon Linux 2 or the Amazon Linux AMI, see Managing Software on Your Linux Instance in the Amazon EC2 User Guide for Linux Instances. For more information about updating your Windows instance, see Updating Your Windows Instance.

Storage

• Use separate Amazon EBS volumes for the operating system versus your data. Ensure that the volume with your data persists after instance termination. For more information, see Preserving Amazon EBS Volumes on Instance Termination (p. 392).

• Use the instance store available for your instance to store temporary data. Remember that the data stored in instance store is deleted when you stop or terminate your instance. If you use instance store for database storage, ensure that you have a cluster with a replication factor that ensures fault tolerance.

Resource Management

• Use instance metadata and custom resource tags to track and identify your AWS resources. For more information, see Instance Metadata and User Data (p. 478) and Tagging Your Amazon EC2 Resources (p. 964).

• View your current limits for Amazon EC2. Plan to request any limit increases in advance of the time that you'll need them. For more information, see Amazon EC2 Service Limits (p. 974).

Backup and Recovery

• Regularly back up your EBS volumes using Amazon EBS snapshots (p. 818), and create an Amazon Machine Image (AMI) (p. 49) from your instance to save the configuration as a template for launching future instances.

• Deploy critical components of your application across multiple Availability Zones, and replicate your data appropriately.

• Design your applications to handle dynamic IP addressing when your instance restarts. For more information, see Amazon EC2 Instance IP Addressing (p. 656).

• Monitor and respond to events. For more information, see Monitoring Amazon EC2 (p. 532).

• Ensure that you are prepared to handle failover. For a basic solution, you can manually attach a network interface or Elastic IP address to a replacement instance. For more information, see Elastic Network Interfaces (p. 679). For an automated solution, you can use Amazon EC2 Auto Scaling. For more information, see the Amazon EC2 Auto Scaling User Guide.

• Regularly test the process of recovering your instances and Amazon EBS volumes if they fail.
Tutorials for Amazon EC2 Instances Running Windows Server

The following tutorials show you how to perform common tasks using EC2 instances running Windows Server.

Tutorials

• Tutorial: Deploying a WordPress Blog on Your Amazon EC2 Instance Running Windows Server (p. 27)
• Tutorial: Installing a WAMP Server on an Amazon EC2 Instance Running Windows Server (p. 31)
• Tutorial: Installing a WIMP Server on an Amazon EC2 Instance Running Windows Server (p. 34)
• Tutorial: Increase the Availability of Your Application on Amazon EC2 (p. 37)
• Tutorial: Setting Up a Windows HPC Cluster on Amazon EC2 (p. 40)

Tutorial: Deploying a WordPress Blog on Your Amazon EC2 Instance Running Windows Server

This tutorial will help you install and deploy a WordPress blog on an Amazon EC2 instance running Windows Server.

If you'd prefer to host your WordPress blog on a Linux instance, see Tutorial: Hosting a WordPress Blog with Amazon EC2 in the Amazon EC2 User Guide for Linux Instances. If you need a high-availability solution with a decoupled database, see Deploying a High-Availability WordPress Website in the AWS Elastic Beanstalk Developer Guide.

Prerequisites

Before you get started, be sure that you do the following:

• Launch an Amazon EC2 instance from a Windows Server AMI. For information, see Getting Started with Amazon EC2 Windows Instances (p. 20).
• Use the AWS free usage tier (if eligible) to launch and use the free Windows t2.micro instance for 12 months. You can use the AWS free usage tier for launching new applications, testing existing applications, or simply gaining hands-on experience with AWS. For more information about eligibility and the highlights, see the AWS Free Usage Tier product page.

Important
If you've launched a regular instance and use it to deploy the WordPress website, you will incur the standard Amazon EC2 usage fees for the instance until you terminate it. For more information about Amazon EC2 usage rates, go to the Amazon EC2 product page.

• Ensure that the security group in which you're launching your instance has ports 80 (HTTP), 443 (HTTPS), and 3389 (RDP) open for inbound traffic. Ports 80 and 443 allow computers outside of the instance to connect with HTTP and HTTPS. If these ports are not open, the WordPress site can't be accessed from outside the instance. Port 3389 allows you to connect to the instance with Remote Desktop Protocol.
• Connect to your instance.
Installing the Microsoft Web Platform Installer

You can use the Microsoft Web Platform Installer to install and configure WordPress on your server. This tool simplifies deployment of Web applications and Web sites to IIS servers. For more information, see Microsoft Web Platform Installer.

To install Microsoft Web Platform Installer

1. Verify that you've met the conditions in Prerequisites (p. 27).
2. Connect to your instance.
3. Disable Internet Explorer Enhanced Security Configuration so that you can download and install required software from the web.
   a. Open Server Manager.
      • On Windows Server 2008 R2, under Server Summary, in the Security Information section, click Configure IE ESC.
      • On Windows Server 2012 R2, click Local Server in the left pane. In the Properties pane, locate IE Enhanced Security Configuration. Click On.
   b. Under Administrators, click Off, and then click OK.
   c. Close Server Manager.
   d. Make a note to re-enable Internet Explorer Enhanced Security Configuration when you have finished installing software from the web.
4. Download and install the latest version of the Microsoft Web Platform Installer.

Installing WordPress

Now you'll use the Web Platform Installer to deploy WordPress on your server.

To install WordPress

1. Download and install Visual C++ Redistributable for Visual Studio 2012 Update 4 or later.
   Important
   Even if your operating system is a 64-bit operating system, you must install the 32-bit version as PHP will not run in 64 bit.
   Open the Web Platform Installer and click Applications.
2. Select WordPress, click Add, and then click Install.
3. On the Prerequisites page, select MySQL for the database to use. Enter the desired administrator password for your MySQL database in the Password and Re-type Password boxes, and then click Continue.
   For more information about creating a secure password, see https://identitysafe.norton.com/password-generator/. Do not reuse an existing password, and make sure to store this password in a safe place.
4. Click I Accept for the list of third-party application software, Microsoft products (including the IIS web server), and components. After the Web Platform Installer finishes installing the software, you are prompted to configure your new site.
5. On the Configure page, clear the default application name in the 'WordPress' application name: box and leave it blank, then leave the default information in the other boxes and click Continue.
6. Click Yes to accept that the contents of the folder will be overwritten.
Configuring Security Keys

WordPress allows you to generate and enter unique authentication keys and salts for your site. These key and salt values provide a layer of encryption to the browser cookies that WordPress users store on their local machines. Basically, adding long, random values here makes your site more secure.

For more information about security keys, see http://codex.wordpress.org/Editing_wp-config.php#Security_Keys.

To configure security keys

1. Visit https://api.wordpress.org/secret-key/1.1/salt/ to randomly generate a set of key values that you can copy and paste into the installation wizard. The following steps will show you how to modify these values in Notepad to work with a Windows installation.

2. Copy all of the text in that page to your clipboard. It should look similar to the example below.

```
define('AUTH_KEY',         '3US+*[RXN8:b~L 0(WU_+ c+WFkI~c)o]~bHw+)/
define('SECURE_AUTH_KEY',  '+{$nDg}');
define('LOGGED_IN_KEY',    '7N`YB3');
define('NONCE_KEY',        '+'9xUahg@::Cj');
define('AUTH_SALT',        '%JT+h');
define('SECURE_AUTH_SALT', '+d!uR+q+{#Z9uFPG.#+S(n~1M&@=gL>U+Nv<zPd@2~Es7Q0-hp28EKy');
define('LOGGED_IN_SALT',   '+'0P/');
define('NONCE_SALT',       '+-97r*k/cgzLmp?Zy4zU4r99qq-2LdP/;)
```

3. Open a Notepad window by clicking Start, All Programs, Accessories, and then Notepad.
4. Paste the copied text into the Notepad window.
5. Windows WordPress installations do not accept the dollar sign ($) in key and salt values, so they need to be replaced with another character (such as S). In the Notepad window, click Edit, then click Replace.
6. In the Find what box, type $.
7. In the Replace with box, type S.
8. Click Replace All to replace all of the dollar signs with S characters.
9. Close the Replace window.
10. Paste the modified key and salt values from the Notepad window into their corresponding boxes in the installation wizard. For example, the AUTH_KEY value in the Notepad window should be pasted into the Authentication Key box in the wizard.

Do not include the single quotes or other text surrounding the values, just the actual value as in the example shown below.

The modified AUTH_KEY line from the Notepad window:

```
define('AUTH_KEY',         '3US+*[RXN8:b~L 0(WU_+ c+WFkI~c)o]~bHw+)/
```

29
Configuring the Site Title and Administrator

When you complete the Web Platform Installer wizard, a browser window opens to your WordPress installation at http://localhost/wp-admin/install.php. On this page, you configure the title for your site and an administrative user to moderate your blog.

To complete the installation

1. On the WordPress Welcome page, enter the following information and click Install WordPress.

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site Title</td>
<td>Enter a name for your WordPress site.</td>
</tr>
<tr>
<td>Username</td>
<td>Enter a name for your WordPress administrator. For security purposes you should choose a unique name for this user, because this will be more difficult to exploit than the default user name, admin.</td>
</tr>
<tr>
<td>Password</td>
<td>Enter a strong password, and then enter it again to confirm. Do not reuse an existing password, and make sure to store this password in a safe place.</td>
</tr>
<tr>
<td>Your E-mail</td>
<td>Enter the email address you want to use for notifications.</td>
</tr>
<tr>
<td>Privacy</td>
<td>Check to allow search engines to index your site.</td>
</tr>
</tbody>
</table>

2. Click Log In.
3. On the Log In page, enter your user name for Username and the site password you entered previously for Password.

Making Your WordPress Site Public

Now that you can see your WordPress blog on your local host, you can publish this website as the default site on your instance so that other people can see it. The next procedure walks you through the process of modifying your WordPress settings to point to the public DNS name of your instance instead of your local host.

To configure the default settings for your WordPress site

1. Open the WordPress dashboard by opening a browser on your instance and going to http://localhost/wp-admin. If prompted for your credentials, enter your user name for the Username and your site password for Password.
2. In the Dashboard pane, click Settings.
3. On the **General Settings** page, enter the following information and click **Save Changes**.

   - **WordPress address (URL)**—The public DNS address of your instance. For example, your URL may look something like http://ec2-203-0-113-25.compute-1.amazonaws.com.

     You can get the public DNS for your instance using the Amazon EC2 console (select the instance and check the **Public DNS** column; if this column is hidden, click the **Show/Hide** icon and select **Public DNS**).

   - **Site address (URL)**—The same public DNS address of your instance that you set in **WordPress address (URL)**.

4. To see your new site, open a browser on a computer other than the instance hosting WordPress and type the public DNS address of your instance in the web address field. Your WordPress site appears.

Congratulations! You have just deployed a WordPress site on a Windows instance.

**Next Steps**

If you no longer need this instance, you can remove it to avoid incurring charges. For more information, see **Clean Up Your Instance** (p. 23).

If your WordPress blog becomes popular and you need more compute power or storage, consider the following steps:

   - Expand the storage space on your instance. For more information, see **Amazon EBS Elastic Volumes** (p. 855).
   - Move your MySQL database to **Amazon RDS** to take advantage of the service's ability to scale automatically.
   - Migrate to a larger instance type. For more information, see **Changing the Instance Type** (p. 188).
   - Add additional instances. For more information, see **Tutorial: Increase the Availability of Your Application on Amazon EC2** (p. 37).


**Tutorial: Installing a WAMP Server on an Amazon EC2 Instance Running Windows Server**

This tutorial shows you how to install an Apache web server with PHP and MySQL on an EC2 instance running Windows Server. This software configuration is sometimes called a WAMP server or WAMP stack (Windows, Apache, MySQL, PHP). For information about how to create a similar server on Linux, see **Tutorial: Installing a LAMP Web Server** in the **Amazon EC2 User Guide for Linux Instances**.

A WAMP stack is designed for easy installation to help developers get up and running quickly. It is not designed for production environments for the following reasons:

   - The default configurations do not meet security requirements for most production environments.
   - Upgrading and patching the different software components on a single production server would affect server availability.
• The WAMP one-click installers do not place files in standard locations, which can make it difficult to locate important configuration files.

You can, however, create a WAMP stack on an EC2 instance to prototype a web project in a controlled test environment. For example, you can host a static website or deploy a dynamic PHP application that reads and writes information to a database.

There are many third-party solutions that you can use to install a WAMP stack; this tutorial uses the Bitnami WAMP stack. For more information, see Review: WAMP stacks for Web developers.

Prerequisites

• Provision a Windows Server 2008 R2 or 2012 R2 base instance. You must configure the base instance with a public domain name system (DNS) name that is reachable from the Internet. For more information, see Getting Started with Amazon EC2 Windows Instances (p. 20).

• Verify that the security group for your instance has the following ports open:
  • Port 80 (HTTP inbound and outbound) - Allows computers outside of the instance to connect by using HTTP.
  • Port 443 (HTTPS inbound and outbound) - Allows computers outside of the instance to connect by using HTTPS.
  • Port 3389 (RDP inbound only) - Allows you to connect to the instance using Remote Desktop Protocol (RDP). As a security best practice, restrict RDP access to a range of IP addresses in your organization.

To install a WAMP server

1. Connect to your instance using Microsoft Remote Desktop. For more information, see Connecting to Your Windows Instance (p. 379).

2. Disable Internet Explorer Enhanced Security Configuration so that you can download and install required software from the web.
   a. From the instance, open Server Manager.
   c. Under Administrators, click Off, and then click OK.
   d. Close Server Manager.
   e. Make a note to re-enable Internet Explorer Enhanced Security Configuration when you have finished installing software from the web.

3. Install software updates to ensure that the instance has the latest security updates and bug fixes.
   • EC2Config - Download and install the latest version of the EC2Config service. For more information, see Installing the Latest Version of EC2Config (p. 406).
   • Windows Update - Run Windows Update to ensure that the latest security and software updates are installed on the instance. In Control Panel, click System and Security. In the Windows Update section, click Check for updates.

4. Download and install the WAMP stack. For the purposes of this tutorial, we suggest that you download and install this WAMP stack. You can, however, download and install other Bitnami WAMP stacks. Regardless of which stack you install, the Bitnami site prompts you to either create a free Bitnami account or log in by using a social media account. After you log in, run the Bitnami setup wizard.
5. After setup completes, verify that the Apache web server is configured properly and running by browsing to a test page. Open a web browser on a different computer and enter either the public DNS address of the WAMP server or the public IP address. The public DNS address for your instance is listed on the Amazon EC2 console in the **Public DNS** column. If this column is hidden, click the **Show/Hide** icon and select **Public DNS**.

**Important**
If you do not see the Bitnami test page, use Windows Firewall with Advanced Security to create a custom rule that allows the HTTP protocol through port 80 and the HTTPS protocol through port 443. For more information, see [Network Security on Microsoft TechNet](https://docs.microsoft.com/en-us/windows/security/network-security/network-security). Also verify that the security group for your instances contains a rule to allow connections on HTTP (port 80). For more information, see [Adding Rules to a Security Group](https://docs.aws.amazon.com/efs/latest/ug/security-groups.html) (p. 588).

6. Test your WAMP server by viewing a PHP file from the web. You must be logged onto the instance as an administrator to perform the following steps.

   a. Create a file named `phpinfo.php` containing the code below and place this file in the Apache root directory. By default, the path is: `C:\Bitnami\wampstack-<version_number>\apache2\htdocs`.

   ```php
   <?php phpinfo(); ?>
   ```

   b. In a web browser, enter the URL of the file you just created. This URL is the public DNS address of your instance followed by a forward slash and the file name. For example: `http://my.public.dns.amazonaws.com/phpinfo.php`.

   c. Verify that the PHP information page is displayed. If the page does not display, verify that you entered the correct public DNS address. Also verify that Windows folder options are configured to show known file extensions. By default, folder options hide known file extensions. If you created the file in Notepad and saved it in the root directory your phpinfo.php file might incorrectly be saved as phpinfo.php.txt.

   d. As a security best practice, delete the phpinfo.php file when you finish testing the WAMP server.

7. Enhance MySQL security by disabling default features and by setting a root password. The `mysql_secure_installation` Perl script can perform these tasks for you. To run the script, you must install Perl.

   a. Download and install Perl from the [Perl Programming Language website](https://www.perl.org/).

   b. In the `C:\Bitnami\wampstack-<version_number>\mysql\bin` directory, double-click `mysql_secure_installation`.

   c. When prompted, enter the MySQL root account password that you entered when you ran the Bitnami WAMP stack installer, and then press Enter.

   d. Type **n** to skip changing the password.

   e. Type **y** to remove the anonymous user accounts.

   f. Type **y** to disable remote root login.

   g. Type **y** to remove the test database.

   h. Type **y** to reload the privilege tables and save your changes.

If you successfully completed the steps in this tutorial, then your WAMP server is functioning properly. To continue testing, you can add more content to the `C:\Bitnami\wampstack-<version_number>\apache2\htdocs` folder and view the content by using the public DNS address for your instance.

**Important**
As a best practice, stop the MySQL server if you do not plan to use it right away. You can restart the server when you need it again.
Tutorial: Installing a WIMP Server on an Amazon EC2 Instance Running Windows Server

This tutorial shows you how to install a Microsoft Internet Information Services (IIS) web server with PHP and MySQL on an EC2 instance running Windows Server. This software configuration is sometimes called a WIMP server or WIMP stack (Windows, IIS, MySQL, PHP).

A WIMP stack is designed for easy installation to help developers get up and running quickly. It is not designed for production environments for the following reasons:

- The default configurations do not meet security requirements for most production environments.
- Upgrading andpatching the different software components on a single production server would affect server availability.
- The WAMP one-click installers do not place files in standard locations, which can make it difficult to locate important configuration files.

You can, however, create a WIMP stack on an EC2 instance to prototype a web project in a controlled test environment. For example, you can host a static website or deploy a dynamic PHP application that reads and writes information to a database.

Prerequisites

- Provision a Windows Server 2008 R2 or 2012 R2 base instance. You must configure the base instance with a public domain name system (DNS) name that is reachable from the Internet. For more information, see Getting Started with Amazon EC2 Windows Instances (p. 20).
- Verify that the security group for your instance has the following ports open:
  - Port 80 (HTTP inbound and outbound) - Allows computers outside of the instance to connect by using HTTP.
  - Port 443 (HTTPS inbound and outbound) - Allows computers outside of the instance to connect by using HTTPS.
  - Port 3389 (RDP inbound only) - Allows you to connect to the instance using Remote Desktop Protocol (RDP). As a security best practice, restrict RDP access to a range of IP addresses in your organization.
- Read the best practices for installing PHP on the Microsoft web platform.

Prepare Your Instance

To prepare your instance

1. Connect to your instance using Microsoft Remote Desktop. For more information, see Connecting to Your Windows Instance (p. 379).
2. Disable Internet Explorer Enhanced Security Configuration so that you can download and install required software from the web.

   Note
   Make a note to re-enable Internet Explorer Enhanced Security Configuration when you have finished installing software from the web.
3. Install software updates to ensure that the instance has the latest security updates and bug fixes.
Install the IIS web server

IIS is a feature of Windows Server and is installed by using Server Manager. The procedure you'll use depends on the version of Windows Server your instance is running.

Install IIS on Windows Server 2012

1. In Server Manager click Add roles and features.
2. On the Before you begin page, click Next.
3. On the Select installation type page, select Role-based or feature-based installation, and then click Next.
4. On the Select destination server page, select your instance from the server pool, and then click Next.
5. On the Select server roles page, select Web Server (IIS), click Add features, and then click Next.
6. On the Select features page, retain the default features and expand .NET Framework 4.5 Features, select ASP.NET 4.5, and then click Next.
7. On the Web Server Role (IIS) page, click Next.
8. On the Select role services page, retain the default services and select Application Development.
9. Expand Application Development, and then select the following features. When selecting these features, if you are prompted, click Add features:
   a. .NET Extensibility 3.5
   b. .NET Extensibility 4.5
   c. Application Initialization
   d. ASP.NET 3.5
   e. ASP.NET 4.5
   f. CGI
10. Click Next.
11. On the Confirm installation selections page, select Restart the destination server automatically if required. When prompted for confirmation, click Yes.
12. Click Install, and then after the installation is complete, click Close.

Install IIS on Windows Server 2008

1. In Server Manager, click Roles.
2. Click Add Roles.
3. On the Before You Begin page, click Next.
5. On the Select Role Services page under Application Development, click ASP.NET.
Install MySQL and PHP

You can download and install MySQL and PHP using the Microsoft Web Platform Installer.

To install MySQL and PHP

1. In your Windows Server instance, download and install the latest version of the Microsoft Web Platform Installer 5.0.
2. In the Microsoft Web Platform Installer, click the Products tab.
3. Select MySQL Windows 5.5 and click Add.
4. Select PHP 5.6.0 and click Add.
5. Click Install.
6. On the Prerequisites page, enter a password for the MySQL default database administrator account, and then click Continue.
7. When the installation is complete, click Finish, and then click Exit to close the Web Platform Installer.

Test Your Server

Test your server by viewing a PHP file from the web. You must be logged onto the instance as an administrator to perform the following steps.

To test your WIMP server

1. Download and install the Visual C++ Redistributable for Visual Studio 2012 Update 4 x86 package. Even if your server is a 64-bit server, you must install the x86 package.
2. Create a file named phpinfo.php that contains the following code and place this file in the IIS root directory. By default, the path is: C:\inetpub\wwwroot.

```php
<?php phpinfo(); ?>
```
3. In a web browser, enter the URL of the file you just created. This URL is the public DNS address of your instance followed by a forward slash and the file name, as in the following example: http://my.public.dns.amazonaws.com/phpinfo.php.

4. Verify that the PHP information page is displayed. If the page does not display, verify that you entered the correct public DNS address. Also verify that Windows folder options are configured to show known file extensions. By default, folder options hide known file extensions. If you created the file in Notepad and saved it in the root directory your phpinfo.php file might incorrectly be saved as phpinfo.php.txt.

5. As a security best practice, delete the phpinfo.php file when you finish testing the WAMP server.

6. Enhance MySQL security by disabling default features and by setting a root password. The mysql_secure_installation Perl script can perform these tasks for you. To run the script, you must install Perl.
   a. Download and install Perl from the Perl Programming Language website.
   b. In the C:\Program Files\MySQL\MySQL Server 5.5\bin directory, double-click mysql_secure_installation.
   c. When prompted, enter the current root password and press Enter.
   d. Type n to skip changing the password.
   e. Type y to remove the anonymous user accounts.
   f. Type y to disable remote root login.
   g. Type y to remove the test database.
   h. Type y to reload the privilege tables and save your changes.

You should now have a fully functional WIMP web server. If you add content to the IIS document root at C:\inetpub\wwwroot, you can view that content at the public DNS address for your instance.

**Important**

As a best practice, stop the MySQL server if you do not plan to use it right away. You can restart the server when you need it again.

---

**Tutorial: Increase the Availability of Your Application on Amazon EC2**

Suppose that you start out running your app or website on a single EC2 instance, and over time, traffic increases to the point that you require more than one instance to meet the demand. You can launch multiple EC2 instances from your AMI and then use Elastic Load Balancing to distribute incoming traffic for your application across these EC2 instances. This increases the availability of your application. Placing your instances in multiple Availability Zones also improves the fault tolerance in your application. If one Availability Zone experiences an outage, traffic is routed to the other Availability Zone.

You can use Amazon EC2 Auto Scaling to maintain a minimum number of running instances for your application at all times. Amazon EC2 Auto Scaling can detect when your instance or application is unhealthy and replace it automatically to maintain the availability of your application. You can also use Amazon EC2 Auto Scaling to scale your Amazon EC2 capacity up or down automatically based on demand, using criteria that you specify.

In this tutorial, we use Amazon EC2 Auto Scaling with Elastic Load Balancing to ensure that you maintain a specified number of healthy EC2 instances behind your load balancer. Note that these instances do not need public IP addresses, because traffic goes to the load balancer and is then routed to the instances. For more information, see Amazon EC2 Auto Scaling and Elastic Load Balancing.
Prerequisites

This tutorial assumes that you have already done the following:

1. Created a virtual private cloud (VPC) with one public subnet in two or more Availability Zones. If you haven't done so, see Create a Virtual Private Cloud (VPC) (p. 16).
2. Launched an instance in the VPC.
3. Connected to the instance and customized it. For example, installing software and applications, copying data, and attaching additional EBS volumes. For information about setting up a web server on your instance, see Tutorial: Installing a WAMP Server on an Amazon EC2 Instance Running Windows Server (p. 31) or Tutorial: Installing a WIMP Server on an Amazon EC2 Instance Running Windows Server (p. 34).
4. Tested your application on your instance to ensure that your instance is configured correctly.
5. Created a custom Amazon Machine Image (AMI) from your instance. For more information, see Creating a Custom Windows AMI (p. 64).
6. (Optional) Terminated the instance if you no longer need it.
7. Created an IAM role that grants your application the access to AWS it needs. For more information, see To create an IAM role using the IAM console (p. 648).

Scale and Load Balance Your Application

Use the following procedure to create a load balancer, create a launch configuration for your instances, create an Auto Scaling group with two or more instances, and associate the load balancer with the Auto Scaling group.

To scale and load-balance your application

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the navigation pane, under LOAD BALANCING, choose Load Balancers.
3. Choose Create Load Balancer.
4. For **Application Load Balancer**, choose **Create**.

5. On the **Configure Load Balancer** page, do the following:
   - a. For **Name**, type a name for your load balancer. For example, `my-lb`.
   - b. For **Scheme**, keep the default value, `internet-facing`.
   - c. For **Listeners**, keep the default, which is a listener that accepts HTTP traffic on port 80.
   - d. For **Availability Zones**, select the VPC that you used for your instances. Select an Availability Zone and then select the public subnet for that Availability Zone. Repeat for a second Availability Zone.
   - e. Choose **Next: Configure Security Settings**.

6. For this tutorial, you are not using a secure listener. Choose **Next: Configure Security Groups**.

7. On the **Configure Security Groups** page, do the following:
   - a. Choose **Create a new security group**.
   - b. Type a name and description for the security group, or keep the default name and description. This new security group contains a rule that allows traffic to the port configured for the listener.
   - c. Choose **Next: Configure Routing**.

8. On the **Configure Routing** page, do the following:
   - a. For **Target group**, keep the default, **New target group**.
   - b. For **Name**, type a name for the target group.
   - c. Keep **Protocol** as HTTP, **Port** as 80, and **Target type** as instance.
   - d. For **Health checks**, keep the default protocol and path.
   - e. Choose **Next: Register Targets**.

9. On the **Register Targets** page, choose **Next: Review** to continue to the next page, as we'll use Amazon EC2 Auto Scaling to add EC2 instances to the target group.

10. On the **Review** page, choose **Create**. After the load balancer is created, choose **Close**.

11. On the navigation pane, under **AUTO SCALING**, choose **Launch Configurations**.
    - • If you are new to Amazon EC2 Auto Scaling, you see a welcome page. Choose **Create Auto Scaling group** to start the Create Auto Scaling Group wizard, and then choose **Create launch configuration**.
    - • Otherwise, choose **Create launch configuration**.

12. On the **Choose AMI** page, select the **My AMIs** tab, and then select the AMI that you created in **Prerequisites** (p. 38).

13. On the **Choose Instance Type** page, select an instance type, and then choose **Next: Configure details**.

14. On the **Configure details** page, do the following:
    - a. For **Name**, type a name for your launch configuration (for example, `my-launch-config`).
    - b. For **IAM role**, select the IAM role that you created in **Prerequisites** (p. 38).
    - c. (Optional) If you need to run a startup script, expand **Advanced Details** and type the script in **User data**.
    - d. Choose **Skip to review**.

15. On the **Review** page, choose **Edit security groups**. You can select an existing security group or create a new one. This security group must allow HTTP traffic and health checks from the load balancer. If your instances will have public IP addresses, you can optionally allow RDP traffic if you need to connect to the instances. When you are finished, choose **Review**.

16. On the **Review** page, choose **Create launch configuration**.

17. When prompted, select an existing key pair, create a new key pair, or proceed without a key pair.
    • Select the acknowledgment check box, and then choose **Create launch configuration**.

39
18. After the launch configuration is created, you must create an Auto Scaling group.
   - If you are new to Amazon EC2 Auto Scaling and you are using the Create Auto Scaling group wizard, you are taken to the next step automatically.
   - Otherwise, choose Create an Auto Scaling group using this launch configuration.

19. On the Configure Auto Scaling group details page, do the following:
   a. For Group name, type a name for the Auto Scaling group. For example, my-asg.
   b. For Group size, type the number of instances (for example, 2). Note that we recommend that you maintain approximately the same number of instances in each Availability Zone.
   c. Select your VPC from Network and your two public subnets from Subnet.
   d. Under Advanced Details, select Receive traffic from one or more load balancers. Select your target group from Target Groups.
   e. Choose Next: Configure scaling policies.

20. On the Configure scaling policies page, choose Review, as we will let Amazon EC2 Auto Scaling maintain the group at the specified size. Note that later on, you can manually scale this Auto Scaling group, configure the group to scale on a schedule, or configure the group to scale based on demand.


22. After the group is created, choose Close.

Test Your Load Balancer

When a client sends a request to your load balancer, the load balancer routes the request to one of its registered instances.

To test your load balancer

1. Verify that your instances are ready. From the Auto Scaling Groups page, select your Auto Scaling group, and then choose the Instances tab. Initially, your instances are in the Pending state. When their states are InService, they are ready for use.
2. Verify that your instances are registered with the load balancer. From the Target Groups page, select your target group, and then choose the Targets tab. If the state of your instances is initial, it’s possible that they are still registering. When the state of your instances is healthy, they are ready for use. After your instances are ready, you can test your load balancer as follows.
3. From the Load Balancers page, select your load balancer.
4. On the Description tab, locate the DNS name. This name has the following form:

   `my-lb-xxxxxxxxxx.us-west-2.elb.amazonaws.com`

5. In a web browser, paste the DNS name for the load balancer into the address bar and press Enter. You’ll see your website displayed.

Tutorial: Setting Up a Windows HPC Cluster on Amazon EC2

You can launch a scalable Windows High Performance Computing (HPC) cluster using Amazon EC2 instances. A Windows HPC cluster requires an Active Directory domain controller, a DNS server, a head node, and one or more compute nodes.

To set up a Windows HPC cluster on Amazon EC2, complete the following tasks:
Prerequisites

You must launch your instances in a VPC. You can use the default VPC or create a nondefault VPC. For more information, see Getting Started in the Amazon VPC User Guide.

Step 1: Create Security Groups

Use the Tools for Windows PowerShell to create security groups for the domain controller, domain members, and the HPC cluster.

To create the security groups

1. Use the $new-EC2SecurityGroup cmdlet to create the security group for the domain controller. Note the ID of the security group in the output.

```powershell
PS C:\> New-EC2SecurityGroup -VpcId vpc-id -GroupName "SG - Domain Controller" -Description "Active Directory Domain Controller"
```

2. Use the $new-EC2SecurityGroup cmdlet to create the security group for the domain members. Note the ID of the security group in the output.

```powershell
PS C:\> New-EC2SecurityGroup -VpcId vpc-id -GroupName "SG - Domain Member" -Description "Active Directory Domain Member"
```

3. Use the $new-EC2SecurityGroup cmdlet to create the security group for the HPC cluster. Note the ID of the security group in the output.

```powershell
PS C:\> New-EC2SecurityGroup -VpcId vpc-id -GroupName "SG - Windows HPC Cluster" -Description "Windows HPC Cluster Nodes"
```

To add rules to the security groups

1. Create the following rules to add to the domain controller security group. Replace the placeholder security group ID with the ID of the domain member security group and the placeholder CIDR block with the CIDR block of your network.

```powershell
PS C:\> $sg_dm = New-Object Amazon.EC2.Model.UserIdGroupPair
PS C:\> $sg_dm.GroupId = "sg-12345678"
PS C:\> $r1 = @{ IpProtocol="UDP"; FromPort="123"; ToPort="123"; UserIdGroupPairs= $sg_dm }
PS C:\> $r2 = @{ IpProtocol="TCP"; FromPort="135"; ToPort="135"; UserIdGroupPairs= $sg_dm }
PS C:\> $r3 = @{ IpProtocol="UDP"; FromPort="138"; ToPort="138"; UserIdGroupPairs= $sg_dm }
PS C:\> $r4 = @{ IpProtocol="TCP"; FromPort="49152"; ToPort="65535"; UserIdGroupPairs= $sg_dm }
```
Step 1: Create Security Groups

1. Create security groups as follows:

   ```powershell
   PS C:\> $r5 = @{ IpProtocol="TCP"; FromPort="389"; ToPort="389"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r6 = @{ IpProtocol="UDP"; FromPort="389"; ToPort="389"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r7 = @{ IpProtocol="TCP"; FromPort="636"; ToPort="636"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r8 = @{ IpProtocol="TCP"; FromPort="3268"; ToPort="3269"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r9 = @{ IpProtocol="TCP"; FromPort="53"; ToPort="53"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r10 = @{ IpProtocol="UDP"; FromPort="53"; ToPort="53"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r11 = @{ IpProtocol="TCP"; FromPort="88"; ToPort="88"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r12 = @{ IpProtocol="UDP"; FromPort="88"; ToPort="88"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r13 = @{ IpProtocol="TCP"; FromPort="445"; ToPort="445"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r14 = @{ IpProtocol="UDP"; FromPort="445"; ToPort="445"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r15 = @{ IpProtocol="ICMP"; FromPort="-1"; ToPort="-1"; UserIdGroupPairs=$sg_dm }
   PS C:\> $r16 = @{ IpProtocol="UDP"; FromPort="53"; ToPort="53"; IpRanges="203.0.113.25/32" }
   PS C:\> $r17 = @{ IpProtocol="TCP"; FromPort="3389"; ToPort="3389"; IpRanges="203.0.113.25/32" }
   ```

2. Use the `Grant-EC2SecurityGroupIngress` cmdlet to add the rules to the domain controller security group.

   ```powershell
   PS C:\> Grant-EC2SecurityGroupIngress -GroupId sg-1a2b3c4d -IpPermission @( $r1, $r2, $r3, $r4, $r5, $r6, $r7, $r8, $r9, $r10, $r11, $r12, $r13, $r14, $r15, $r16, $r17 )
   ```

For more information about these security group rules, see the following Microsoft article: How to configure a firewall for domains and trusts.

3. Create the following rules to add to the domain member security group. Replace the placeholder security group ID with the ID of the domain controller security group.

   ```powershell
   PS C:\> $sg_dc = New-Object Amazon.EC2.Model.UserIdGroupPair
   PS C:\> $sg_dc.GroupId = "sg-1a2b3c4d"
   PS C:\> $r1 = @{ IpProtocol="TCP"; FromPort="49152"; ToPort="65535"; UserIdGroupPairs=$sg_dc }
   PS C:\> $r2 = @{ IpProtocol="UDP"; FromPort="49152"; ToPort="65535"; UserIdGroupPairs=$sg_dc }
   PS C:\> $r3 = @{ IpProtocol="TCP"; FromPort="53"; ToPort="53"; UserIdGroupPairs=$sg_dc }
   PS C:\> $r4 = @{ IpProtocol="UDP"; FromPort="53"; ToPort="53"; UserIdGroupPairs=$sg_dc }
   ```

4. Use the `Grant-EC2SecurityGroupIngress` cmdlet to add the rules to the domain member security group.

   ```powershell
   PS C:\> Grant-EC2SecurityGroupIngress -GroupId sg-12345678 -IpPermission @( $r1, $r2, $r3, $r4 )
   ```

5. Create the following rules to add to the HPC cluster security group. Replace the placeholder security group ID with the ID of the HPC cluster security group and the placeholder CIDR block with the CIDR block of your network.

   ```powershell
   $sg_hpc = New-Object Amazon.EC2.Model.UserIdGroupPair
   PS C:\> $sg_hpc.GroupId = "sg-87654321"
   ```
Step 1: Create Security Groups

6. Use the `Grant-EC2SecurityGroupIngress` cmdlet to add the rules to the HPC cluster security group.

```powershell
PS C:\> Grant-EC2SecurityGroupIngress -GroupId sg-87654321 -IpPermission @(
    $r1, $r2,
    $r3, $r4, $r5, $r6, $r7, $r8, $r9, $r10, $r11, $r12, $r13, $r14, $r15, $r16, $r17,
    $r18, $r19, $r20, $r21, $r22, $r23, $r24
)
```

For more information about these security group rules, see the following Microsoft article: HPC Cluster Networking: Windows Firewall configuration.

7. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.

8. In the navigation pane, choose Security Groups. Verify that the all three security groups appear in the list and have the required rules.
Step 2: Set Up Your Active Directory Domain Controller

The Active Directory domain controller provides authentication and centralized resource management of the HPC environment and is required for the installation. To set up your Active Directory, launch an instance to serve as the domain controller for your HPC cluster and configure it.

To launch a domain controller for your HPC cluster

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the console dashboard, choose Launch Instance.
3. On the Choose an AMI page, select an AMI for Windows Server, and choose Select.
4. On the next page of the wizard, select an instance type, then choose Next: Configure Instance Details.
5. On the Configure Instance Details page, select your VPC from Network and a subnet from Subnet. On the next page of the wizard, you can specify additional storage for your instance.
6. On the Add Tags page, enter Domain Controller as the value for the Name tag for the instance, and then choose Next: Configure Security Group.
7. On the Configure Security Group page, choose Select an existing security group, choose the SG – Domain Controller security group, and then choose Review and Launch.
8. Choose Launch.
9. In the navigation pane, choose Elastic IPs.
11. Select the Elastic IP address you created, and choose Actions, Associate address. For Instance, choose the domain controller instance. Choose Associate.

Connect to the instance you created, and configure the server as a domain controller for the HPC cluster.

To configure your instance as a domain controller

1. Connect to your Domain Controller instance. For more information, see Connecting to Your Windows Instance (p. 379).
2. Open Server Manager, and add the Active Directory Domain Services role.
3. Promote the server to a domain controller using Server Manager or by running DCPromo.exe.
4. Create a new domain in a new forest.
5. Type hpc.local as the fully qualified domain name (FQDN).
7. Ensure that the DNS Server option is selected, and then choose Next.
8. Select Yes, the computer will use an IP address automatically assigned by a DHCP server (not recommended).
9. When prompted, choose Yes to continue.
10. Complete the wizard and then select Reboot on Completion.
11. Connect to the instance as hpc.local\administrator.
12. Create a domain user hpc.local\hpcuser.

Step 3: Configure Your Head Node

An HPC client connects to the head node. The head node facilitates the scheduled jobs. You configure your head node by launching an instance, installing the HPC Pack, and configuring the cluster.
Launch an instance and then configure it as a member of the hpc.local domain and with the necessary user accounts.

To configure an instance as your head node

1. Launch an instance and name it **HPC-Head**. When you launch the instance, select both of these security groups: SG - Windows HPC Cluster and SG - Domain Member.
2. Connect to the instance and get the existing DNS server address using the following command:

   ```
   ipconfig /all
   ```
3. Update the TCP/IPv4 properties of the **HPC-Head** NIC to include the Elastic IP address for the Domain Controller instance as the primary DNS, and then add the additional DNS IP address from the previous step.
4. Join the machine to the hpc.local domain using the credentials for hpc.local\administrator (the domain administrator account).
5. Add hpc.local\hpcuser as the local administrator. When prompted for credentials, use hpc.local\administrator, and then restart the instance.
6. Connect to **HPC-Head** as hpc.local\hpcuser.

To install the HPC Pack

1. Connect to your **HPC-Head** instance using the hpc.local\hpcuser account.
2. Using **Server Manager**, turn off Internet Explorer Enhanced Security Configuration (IE ESC) for Administrators.
   a. In **Server Manager**, under **Security Information**, choose **Configure IE ESC**.
   b. Turn off IE ESC for administrators.
3. Install the HPC Pack on **HPC-Head**.
   a. Download the HPC Pack to **HPC-Head** from the Microsoft Download Center. Choose the HPC Pack for the version of Windows Server on **HPC-Head**.
   b. Extract the files to a folder, open the folder, and double-click `setup.exe`.
   c. On the Installation page, select **Create a new HPC cluster by creating a head node**, and then choose **Next**.
   d. Accept the default settings to install all the databases on the Head Node, and then choose **Next**.
   e. Complete the wizard.

To configure your HPC cluster on the head node

1. Start **HPC Cluster Manager**.
2. In the **Deployment To-Do List**, select **Configure your network**.
   a. In the wizard, select the default option (5), and then choose **Next**.
   b. Complete the wizard accepting default values on all screens, and choose how you want to update the server and participate in customer feedback.
   c. Choose **Configure**.
3. Select **Provide Network Credentials**, then provide the hpc.local\hpcuser credentials.
4. Select **Configure the naming of new nodes**, and then choose **OK**.
5. Select **Create a node template**.
   a. Select the **Compute node template**, and then choose **Next**.
b. Select **Without operating system**, and then continue with the defaults.

c. Choose **Create**.

### Step 4: Set Up the Compute Node

You set up the compute node by launching an instance, installing the HPC Pack, and adding the node to your cluster.

First, launch an instance, and then configure it as a member of the `hpc.local` domain with the necessary user accounts.

**To configure an instance for your compute node**

1. Launch an instance and name it `HPC-Compute`. When you launch the instance, select the following security groups: **SG - Windows HPC Cluster** and **SG - Domain Member**.
2. Log in to the instance and get the existing DNS server address from `HPC-Compute` using the following command:

   ```cmd
   IPConfig /all
   ```

3. Update the TCP/IPv4 properties of the `HPC-Compute` NIC to include the Elastic IP address of the Domain Controller instance as the primary DNS. Then add the additional DNS IP address from the previous step.
4. Join the machine to the `hpc.local` domain using the credentials for `hpc.local\administrator` (the domain administrator account).
5. Add `hpc.local\hpcuser` as the local administrator. When prompted for credentials, use `hpc.local\administrator`, and then restart.
6. Connect to `HPC-Compute` as `hpc.local\hpcuser`.

**To install the HPC Pack on the compute node**

1. Connect to your `HPC-Compute` instance using the `hpc.local\hpcuser` account.
2. Using **Server Manager**, turn off Internet Explorer Enhanced Security Configuration (IE ESC) for Administrators.
   a. In **Server Manager**, under **Security Information**, choose **Configure IE ESC**.
   b. Turn off IE ESC for administrators.
3. Install the HPC Pack on `HPC-Compute`.
   a. Download the HPC Pack to `HPC-Compute` from the [Microsoft Download Center](https://www.microsoft.com). Choose the HPC Pack for the version of Windows Server on `HPC-Compute`.
   b. Extract the files to a folder, open the folder, and double-click `setup.exe`.
   c. On the **Installation** page, select **Join an existing HPC cluster by creating a new compute node**, and then choose **Next**.
   d. Specify the fully-qualified name of the `HPC-Head` instance, and then choose the defaults.
   e. Complete the wizard.

To complete your cluster configuration, from the head node, add the compute node to your cluster.

**To add the compute node to your cluster**

1. Connect to the `HPC-Head` instance as `hpc.local\hpcuser`. 
2. Open **HPC Cluster Manager**.
3. Select **Node Management**.
4. If the compute node displays in the **Unapproved** bucket, right-click the node that is listed and select **Add Node**.
   a. Select **Add compute nodes or broker nodes that have already been configured**.
   b. Select the check box next to the node and choose **Add**.
5. Right-click the node and choose **Bring Online**.

### Step 5: Scale Your HPC Compute Nodes (Optional)

**To scale your compute nodes**

1. Connect to the **HPC-Compute** instance as **hpc.local\hpcuser**.
2. Delete any files you downloaded locally from the HP Pack installation package. (You have already run setup and created these files on your image so they do not need to be cloned for an AMI.)
3. From **C:\Program Files\Amazon\Ec2ConfigService** open the file **sysprep2008.xml**.
4. At the bottom of `<settings pass="specialize">`, add the following section. Make sure to replace **hpc.local**, **password**, and **hpcuser** to match your environment.

   ```xml
     <Identification>
       <UnsecureJoin>false</UnsecureJoin>
       <Credentials>
         <Domain>hpc.local</Domain>
         <Password>password</Password>
         <Username>hpcuser</Username>
       </Credentials>
       <JoinDomain>hpc.local</JoinDomain>
     </Identification>
   </component>
   ```

5. Save **sysprep2008.xml**.
6. Choose **Start**, **All Programs**, **EC2ConfigService Settings**.
   a. Choose the **General** tab, and clear the **Set Computer Name** check box.
   b. Choose the **Bundle** tab, and then choose **Run Sysprep and Shutdown Now**.
7. Open the Amazon EC2 console.
8. In the navigation pane, choose **Instances**.
9. Wait for the instance status to show **stopped**.
10. Select the instance, choose **Actions**, **Image**, **Create Image**.
11. Specify an image name and image description, and then choose **Create Image** to create an AMI from the instance.
12. Start the original **HPC-Compute** instance that was shut down.
13. Connect to the head node using the **hpc.local\hpcuser** account.
14. From **HPC Cluster Manager**, delete the old node that now appears in an error state.
15. In the Amazon EC2 console, in the navigation pane, choose **AMIs**.
16. Use the AMI you created to add additional nodes to the cluster.
You can launch additional compute nodes from the AMI that you created. These nodes are automatically joined to the domain, but you must add them to the cluster as already configured nodes in HPC Cluster Manager using the head node and then bring them online.
Amazon Machine Images (AMI)

An Amazon Machine Image (AMI) provides the information required to launch an instance. You must specify an AMI when you launch an instance. You can launch multiple instances from a single AMI when you need multiple instances with the same configuration. You can use different AMIs to launch instances when you need instances with different configurations.

An AMI includes the following:

- One or more EBS snapshots, or, for instance-store-backed AMIs, a template for the root volume of the instance (for example, an operating system, an application server, and applications).
- Launch permissions that control which AWS accounts can use the AMI to launch instances.
- A block device mapping that specifies the volumes to attach to the instance when it's launched.

Creating Your Own AMI

You can launch an instance from an existing AMI, customize the instance, and then save this updated configuration as a custom AMI. Instances launched from this new custom AMI include the customizations that you made when you created the AMI.

For more information, see Creating a Custom Windows AMI (p. 64).

To help categorize and manage your AMIs, you can assign custom tags to them. For more information, see Tagging Your Amazon EC2 Resources (p. 964).

Buying, Sharing, and Selling AMIs

After you create an AMI, you can keep it private so that only you can use it, or you can share it with a specified list of AWS accounts. You can also make your custom AMI public so that the community can use it. Building a safe, secure, usable AMI for public consumption is a fairly straightforward process, if you follow a few simple guidelines. For information about how to create and use shared AMIs, see Shared AMIs (p. 53).

You can purchase AMIs from a third party, including AMIs that come with service contracts from organizations such as Red Hat. You can also create an AMI and sell it to other Amazon EC2 users. For more information about buying or selling AMIs, see Paid AMIs (p. 60).

Deregistering Your AMI

You can deregister an AMI when you have finished with it. After you deregister an AMI, it can't be used to launch new instances. Existing instances launched from the AMI are not affected. For more information, see Deregistering Your Windows AMI (p. 77).

AWS Windows AMIs

AWS provides a set of publicly available AMIs that contain software configurations specific to the Windows platform. Using these AMIs, you can quickly start building and deploying your applications.
using Amazon EC2. First choose the AMI that meets your specific requirements, and then launch an instance using that AMI. You retrieve the password for the administrator account and then log in to the instance using Remote Desktop Connection, just as you would with any other Windows server.

Some Windows AMIs include an edition of Microsoft SQL Server (SQL Enterprise Edition, SQL Server Standard, SQL Server Express, or SQL Server Web). Launching an instance from a Windows AMI with Microsoft SQL Server enables you to run the instance as a database server. Alternatively, you can launch an instance from any Windows AMI and then install the database software that you need on the instance.

Microsoft no longer supports Windows Server 2003 (see Microsoft Windows Server 2003 End-of-Support). We recommend that you launch new EC2 instances using a supported version of Windows Server. If you have existing EC2 instances that are running an unsupported version of Windows Server, we recommend that you upgrade those instances to a supported version of Windows Server. For more information, see Upgrading an Amazon EC2 Windows Instance to a Newer Version of Windows Server (p. 495).

## Selecting an Initial Windows AMI

To view the Windows AMIs provided by AWS, you can use the Amazon EC2 console or AWS Marketplace. For more information, see Finding a Windows AMI (p. 51).

You can also create an AMI from your own Windows computer. For more information, see the following services:

- AWS Server Migration Service
- VM Import/Export

## Keeping Your AMIs Up-to-Date

AWS provides updated, fully-patched Windows AMIs within five business days of Microsoft's patch Tuesday (the second Tuesday of each month). For more information, see Details About AWS Windows AMI Versions (p. 82).

The AWS Windows AMIs contain the latest security updates available at the time they were created. For more information, see Updating Your Windows Instance (p. 79).

## Virtualization Types

AMIs use one of two types of virtualization: paravirtual (PV) or hardware virtual machine (HVM). The main differences between PV and HVM AMIs are the way in which they boot and whether they can take advantage of special hardware extensions for better performance. Windows AMIs are HVM AMIs.

HVM AMIs are presented with a fully virtualized set of hardware and boot by executing the master boot record of the root block device of your image. This virtualization type provides the ability to run an operating system directly on top of a virtual machine without any modification, as if it were run on the bare-metal hardware. The Amazon EC2 host system emulates some or all of the underlying hardware that is presented to the guest.

HVM guests can take advantage of hardware extensions that provide fast access to the underlying hardware on the host system. HVM AMIs are required to take advantage of enhanced networking and GPU processing. In order to pass through instructions to specialized network and GPU devices, the OS needs to be able to have access to the native hardware platform; HVM virtualization provides this access.

Paravirtual guests traditionally performed better with storage and network operations than HVM guests because they could leverage special drivers for I/O that avoided the overhead of emulating network
and disk hardware, whereas HVM guests had to translate these instructions to emulated hardware. Now PV drivers are available for HVM guests, so Windows instances can get performance advantages in storage and network I/O by using them. With these PV on HVM drivers, HVM guests can get the same performance as paravirtual guests, or better.

Finding a Windows AMI

Before you can launch an instance, you must select an AMI to use. As you select an AMI, consider the following requirements you might have for the instances that you’ll launch:

- The Region
- The operating system
- The architecture: 32-bit (i386), 64-bit (x86_64), or 64-bit ARM (arm64)
- The provider (for example, Amazon Web Services)
- Additional software (for example, SQL server)

If you need to find a Linux AMI, see Finding a Linux AMI in the Amazon EC2 User Guide for Linux Instances.

Contents

- Finding a Windows AMI Using the Amazon EC2 Console (p. 51)
- Finding an AMI Using the AWS Tools for Windows PowerShell (p. 52)
- Finding an AMI Using the AWS CLI (p. 52)

Finding a Windows AMI Using the Amazon EC2 Console

You can find Windows AMIs using the Amazon EC2 console. You can search through all available AMIs using the Images page, or select from commonly used AMIs on the Quick Start tab when you use the console to launch an instance. AMI IDs are unique to each region.

To find a Windows AMI using the Choose AMI page

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region in which to launch your instances. You can select any Region that's available to you, regardless of your location.
3. From the console dashboard, choose Launch Instance.
4. On the Quick Start tab, select from one of the commonly used AMIs in the list. If you don’t see the AMI that you need, select the AWS Marketplace or Community AMIs tab to find additional AMIs.

To find a Windows AMI using the Images page

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region in which to launch your instances. You can select any Region that's available to you, regardless of your location.
3. In the navigation pane, choose AMIs.
4. (Optional) Use the Filter options to scope the list of displayed AMIs to see only the AMIs that interest you. For example, to list all Windows AMIs provided by AWS, select Public images. Choose
Finding an AMI Using the AWS Tools for Windows PowerShell

You can use cmdlets for Amazon EC2 or AWS Systems Manager to list only the Windows AMIs that meet your needs. After locating an AMI that meets your needs, make note of its ID so that you can use it to launch instances. For more information, see Launch an Instance Using Windows PowerShell in the AWS Tools for Windows PowerShell User Guide.

Amazon EC2

For information and examples, see Find an AMI Using Windows PowerShell in the AWS Tools for Windows PowerShell User Guide.

Systems Manager Parameter Store

For information and examples, see Query for the Latest Windows AMI Using Systems Manager Parameter Store.

Finding an AMI Using the AWS CLI

You can use AWS CLI commands for Amazon EC2 or AWS Systems Manager to list only the Windows AMIs that meet your needs. After locating an AMI that meets your needs, make note of its ID so that you can use it to launch instances. For more information, see Launching an Instance Using the AWS CLI in the AWS Command Line Interface User Guide.

Amazon EC2

The describe-images command supports filtering parameters. For example, use the --owners parameter to display public AMIs owned by Amazon.

```
aws ec2 describe-images --owners self amazon
```

You can add the following filter to the previous command to display only Windows AMIs:

```
--filters "Name=platform,Values=windows"
```

Important

Omitting the --owners flag from the describe-images command will return all images for which you have launch permissions, regardless of ownership.

Systems Manager Parameter Store

For information and examples, see Query for the Latest Windows AMI Using Systems Manager Parameter Store.
Shared AMIs

A shared AMI is an AMI that a developer created and made available for other developers to use. One of the easiest ways to get started with Amazon EC2 is to use a shared AMI that has the components you need and then add custom content. You can also create your own AMIs and share them with others.

You use a shared AMI at your own risk. Amazon can't vouch for the integrity or security of AMIs shared by other Amazon EC2 users. Therefore, you should treat shared AMIs as you would any foreign code that you might consider deploying in your own data center and perform the appropriate due diligence. We recommend that you get an AMI from a trusted source. If you have questions or observations about a shared AMI, use the AWS forums.

Amazon's public images have an aliased owner, which appears as amazon in the account field. This enables you to find AMIs from Amazon easily. Other users can't alias their AMIs.

For information about creating an AMI, see Creating an Amazon EBS-Backed Windows AMI. For more information about building, delivering, and maintaining your applications on the AWS Marketplace, see the AWS Marketplace User Guide and AWS Marketplace Seller Guide.

Contents
- Finding Shared AMIs (p. 53)
- Making an AMI Public (p. 55)
- Sharing an AMI with Specific AWS Accounts (p. 57)
- Using Bookmarks (p. 59)
- Guidelines for Shared Windows AMIs (p. 60)

Finding Shared AMIs

You can use the Amazon EC2 console or the command line to find shared AMIs.

**Note**
AMIs are a regional resource. Therefore, when searching for a shared AMI (public or private), you must search for it from within the region from which it is being shared. To make an AMI available in a different region, copy the AMI to the region and then share it. For more information, see Copying an AMI.

Finding a Shared AMI (Console)

**To find a shared private AMI using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **AMIs**.
3. In the first filter, choose **Private images**. All AMIs that have been shared with you are listed. To granulate your search, choose the Search bar and use the filter options provided in the menu.

**To find a shared public AMI using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **AMIs**.
3. In the first filter, choose **Public images**. To granulate your search, choose the Search bar and use the filter options provided in the menu.
4. Use filters to list only the types of AMIs that interest you. For example, choose **Owner** and then choose **Amazon images** to display only Amazon's public images.

**Finding a Shared AMI (Tools for Windows PowerShell)**

Use the `Get-EC2Image` command (Tools for Windows PowerShell) to list AMIs. You can scope the list to the types of AMIs that interest you, as shown in the following examples.

**Example: List all public AMIs**

The following command lists all public AMIs, including any public AMIs that you own.

```
PS C:\> Get-EC2Image -ExecutableUser all
```

**Example: List AMIs with explicit launch permissions**

The following command lists the AMIs for which you have explicit launch permissions. This list does not include any AMIs that you own.

```
PS C:\> Get-EC2Image -ExecutableUser self
```

**Example: List AMIs owned by Amazon**

The following command lists the AMIs owned by Amazon. Amazon's public AMIs have an aliased owner, which appears as `amazon` in the account field. This enables you to find AMIs from Amazon easily. Other users can't alias their AMIs.

```
PS C:\> Get-EC2Image -Owner amazon
```

**Example: List AMIs owned by an account**

The following command lists the AMIs owned by the specified AWS account.

```
PS C:\> Get-EC2Image -Owner 123456789012
```

**Example: Scope AMIs using a filter**

To reduce the number of displayed AMIs, use a filter to list only the types of AMIs that interest you. For example, use the following filter to display only EBS-backed AMIs.

```
-Filter @{ Name="root-device-type"; Values="ebs" }
```

**Finding a Shared AMI (AWS CLI)**

Use the `describe-images` command (AWS CLI) to list AMIs. You can scope the list to the types of AMIs that interest you, as shown in the following examples.

**Example: List all public AMIs**

The following command lists all public AMIs, including any public AMIs that you own.

```
aws ec2 describe-images --executable-users all
```
Example: List AMIs with explicit launch permissions

The following command lists the AMIs for which you have explicit launch permissions. This list does not include any AMIs that you own.

```
aws ec2 describe-images --executable-users self
```

Example: List AMIs owned by Amazon

The following command lists the AMIs owned by Amazon. Amazon's public AMIs have an aliased owner, which appears as `amazon` in the account field. This enables you to find AMIs from Amazon easily. Other users can't alias their AMIs.

```
aws ec2 describe-images --owners amazon
```

Example: List AMIs owned by an account

The following command lists the AMIs owned by the specified AWS account.

```
aws ec2 describe-images --owners 123456789012
```

Example: Scope AMIs using a filter

To reduce the number of displayed AMIs, use a filter to list only the types of AMIs that interest you. For example, use the following filter to display only EBS-backed AMIs.

```
--filters "Name=root-device-type,Values=ebs"
```

Making an AMI Public

Amazon EC2 enables you to share your AMIs with other AWS accounts. You can allow all AWS accounts to launch the AMI (make the AMI public), or only allow a few specific accounts to launch the AMI (see Sharing an AMI with Specific AWS Accounts (p. 57)). You are not billed when your AMI is launched by other AWS accounts; only the accounts launching the AMI are billed.

AMIs with encrypted volumes cannot be made public.

AMIs are a regional resource. Therefore, sharing an AMI makes it available in that region. To make an AMI available in a different region, copy the AMI to the region and then share it. For more information, see Copying an AMI (p. 72).

**Note**

If an AMI has a product code, or contains a snapshot of an encrypted volume, you can't make it public. You can share the AMI only with specific AWS accounts.

Sharing an AMI with all AWS Accounts (Console)

After you make an AMI public, it is available in **Community AMIs** when you launch an instance in the same region using the console. Note that it can take a short while for an AMI to appear in **Community AMIs** after you make it public. It can also take a short while for an AMI to be removed from **Community AMIs** after you make it private again.

**To share a public AMI using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **AMIs**.
3. Select your AMI from the list, and then choose **Actions, Modify Image Permissions**.
4. Choose **Public** and choose **Save**.

### Sharing an AMI with all AWS Accounts (Tools for Windows PowerShell)

Each AMI has a `launchPermission` property that controls which AWS accounts, besides the owner's, are allowed to use that AMI to launch instances. By modifying the `launchPermission` property of an AMI, you can make the AMI public (which grants launch permissions to all AWS accounts) or share it with only the AWS accounts that you specify.

You can add or remove account IDs from the list of accounts that have launch permissions for an AMI. To make the AMI public, specify the `all` group. You can specify both public and explicit launch permissions.

**To make an AMI public**

1. Use the `Edit-EC2ImageAttribute` command as follows to add the `all` group to the `launchPermission` list for the specified AMI.
   ```bash
   PS C:\> Edit-EC2ImageAttribute -ImageId ami-0abcdef1234567890 -Attribute launchPermission -OperationType add -UserGroup all
   ```
2. To verify the launch permissions of the AMI, use the following `Get-EC2ImageAttribute` command.
   ```bash
   PS C:\> Get-EC2ImageAttribute -ImageId ami-0abcdef1234567890 -Attribute launchPermission
   ```
3. (Optional) To make the AMI private again, remove the `all` group from its launch permissions. Note that the owner of the AMI always has launch permissions and is therefore unaffected by this command.
   ```bash
   PS C:\> Edit-EC2ImageAttribute -ImageId ami-0abcdef1234567890 -Attribute launchPermission -OperationType remove -UserGroup all
   ```

### Sharing an AMI with all AWS Accounts (AWS CLI)

Each AMI has a `launchPermission` property that controls which AWS accounts, besides the owner's, are allowed to use that AMI to launch instances. By modifying the `launchPermission` property of an AMI, you can make the AMI public (which grants launch permissions to all AWS accounts) or share it with only the AWS accounts that you specify.

You can add or remove account IDs from the list of accounts that have launch permissions for an AMI. To make the AMI public, specify the `all` group. You can specify both public and explicit launch permissions.

**To make an AMI public**

1. Use the `modify-image-attribute` command as follows to add the `all` group to the `launchPermission` list for the specified AMI.
   ```bash
   aws ec2 modify-image-attribute --image-id ami-0abcdef1234567890 --launch-permission "Add=[{Group=all}]"
   ```
2. To verify the launch permissions of the AMI, use the following `describe-image-attribute` command.
Sharing an AMI with Specific AWS Accounts

You can share an AMI with specific AWS accounts without making the AMI public. All you need is the AWS account IDs. If you share an AMI with encrypted volumes, you must also share any CMKs used to encrypt them. For more information, see Sharing an Amazon EBS Snapshot (p. 842).

AMIs are a regional resource. Therefore, sharing an AMI makes it available in that region. To make an AMI available in a different region, copy the AMI to the region and then share it. For more information, see Copying an AMI (p. 72).

There is no limit to the number of AWS accounts with which an AMI can be shared.

Sharing an AMI (Console)

To grant explicit launch permissions using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose AMIs.
3. Select your AMI in the list, and then choose Actions, Modify Image Permissions.
4. Specify the AWS account number of the user with whom you want to share the AMI in the AWS Account Number field, then choose Add Permission.

   To share this AMI with multiple users, repeat this step until you have added all the required users.
5. To allow create volume permissions for snapshots, select Add "create volume" permissions to the following associated snapshots when creating permissions.

   Note
   You do not need to share the Amazon EBS snapshots that an AMI references in order to share the AMI. Only the AMI itself needs to be shared; the system automatically provides the instance access to the referenced Amazon EBS snapshots for the launch. However, you do need to share any CMKs used to encrypt snapshots that the AMI references. For more information, see Sharing an Amazon EBS Snapshot (p. 842).
6. Choose Save when you are done.
7. (Optional) To view the AWS account IDs with which you have shared the AMI, select the AMI in the list, and choose the Permissions tab. To find AMIs that are shared with you, see Finding Shared AMIs (p. 53).

Sharing an AMI (Tools for Windows PowerShell)

Use the Edit-EC2ImageAttribute command (Tools for Windows PowerShell) to share an AMI as shown in the following examples.
To grant explicit launch permissions

The following command grants launch permissions for the specified AMI to the specified AWS account.

```
PS C:\> Edit-EC2ImageAttribute -ImageId ami-0abcdef1234567890 -Attribute launchPermission -OperationType add -UserId "123456789012"
```

The following command grants create volume permission for a snapshot.

```
PS C:\> Edit-EC2SnapshotAttribute -SnapshotId snap-1234567890abcdef0 -Attribute CreateVolumePermission -OperationType add -UserId 123456789012
```

**Note**
You do not need to share the Amazon EBS snapshots that an AMI references in order to share the AMI. Only the AMI itself needs to be shared; the system automatically provides the instance access to the referenced Amazon EBS snapshots for the launch. However, you do need to share any CMKs used to encrypt snapshots that the AMI references. For more information, see Sharing an Amazon EBS Snapshot (p. 842).

To remove launch permissions for an account

The following command removes launch permissions for the specified AMI from the specified AWS account:

```
PS C:\> Edit-EC2ImageAttribute -ImageId ami-0abcdef1234567890 -Attribute launchPermission -OperationType remove -UserId "123456789012"
```

The following command removes create volume permission for a snapshot.

```
PS C:\> Edit-EC2SnapshotAttribute -SnapshotId snap-1234567890abcdef0 -Attribute CreateVolumePermission -OperationType remove -UserId 123456789012
```

To remove all launch permissions

The following command removes all public and explicit launch permissions from the specified AMI. Note that the owner of the AMI always has launch permissions and is therefore unaffected by this command.

```
PS C:\> Reset-EC2ImageAttribute -ImageId ami-0abcdef1234567890 -Attribute launchPermission
```

Sharing an AMI (AWS CLI)

Use the modify-image-attribute command (AWS CLI) to share an AMI as shown in the following examples.

**To grant explicit launch permissions**

The following command grants launch permissions for the specified AMI to the specified AWS account.

```
aws ec2 modify-image-attribute --image-id ami-0abcdef1234567890 --launch-permission "Add=[{UserId=123456789012}]"
```

The following command grants create volume permission for a snapshot.
### Using Bookmarks

If you have created a public AMI, or shared an AMI with another AWS user, you can create a **bookmark** that allows a user to access your AMI and launch an instance in their own account immediately. This is an easy way to share AMI references, so users don’t have to spend time finding your AMI in order to use it.

Note that your AMI must be public, or you must have shared it with the user to whom you want to send the bookmark.

#### To create a bookmark for your AMI

1. Type a URL with the following information, where `region` is the region in which your AMI resides:

   ```
   ```

   For example, this URL launches an instance from the ami-0abcdef1234567890 AMI in the us-east-1 region:

   ```
   https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#LaunchInstanceWizard:ami=ami-0abcdef1234567890
   ```

2. Distribute the link to users who want to use your AMI.
3. To use a bookmark, choose the link or copy and paste it into your browser. The launch wizard opens, with the AMI already selected.

Guidelines for Shared Windows AMIs

Use the following guidelines to reduce the attack surface and improve the reliability of the AMIs you create.

- No list of security guidelines can be exhaustive. Build your shared AMIs carefully and take time to consider where you might expose sensitive data.
- Develop a repeatable process for building, updating, and republishing AMIs.
- Build AMIs using the most up-to-date operating systems, packages, and software.
- Download and install the latest version of the EC2Config service. For more information about installing this service, see Installing the Latest Version of EC2Config (p. 406).
- Verify that Ec2SetPassword, Ec2WindowsActivate and Ec2HandleUserData are enabled.
- Verify that no guest accounts or Remote Desktop user accounts are present.
- Disable or remove unnecessary services and programs to reduce the attack surface of your AMI.
- Remove instance credentials, such as your key pair, from the AMI (if you saved them on the AMI). Store the credentials in a safe location.
- Ensure that the administrator password and passwords on any other accounts are set to an appropriate value for sharing. These passwords are available for anyone who launches your shared AMI.
- Test your AMI before you share it.

Paid AMIs

A paid AMI is an AMI that you can purchase from a developer.

Amazon EC2 integrates with AWS Marketplace, enabling developers to charge other Amazon EC2 users for the use of their AMIs or to provide support for instances.

The AWS Marketplace is an online store where you can buy software that runs on AWS, including AMIs that you can use to launch your EC2 instance. The AWS Marketplace AMIs are organized into categories, such as Developer Tools, to enable you to find products to suit your requirements. For more information about AWS Marketplace, see the AWS Marketplace site.

Launching an instance from a paid AMI is the same as launching an instance from any other AMI. No additional parameters are required. The instance is charged according to the rates set by the owner of the AMI, as well as the standard usage fees for the related web services, for example, the hourly rate for running an m1.small instance type in Amazon EC2. Additional taxes might also apply. The owner of the paid AMI can confirm whether a specific instance was launched using that paid AMI.

Important
Amazon DevPay is no longer accepting new sellers or products. AWS Marketplace is now the single, unified e-commerce platform for selling software and services through AWS. For information about how to deploy and sell software from AWS Marketplace, see Selling on AWS Marketplace. AWS Marketplace supports AMIs backed by Amazon EBS.

Contents
- Selling Your AMI (p. 61)
- Finding a Paid AMI (p. 61)
- Purchasing a Paid AMI (p. 62)
Selling Your AMI

You can sell your AMI using AWS Marketplace. AWS Marketplace offers an organized shopping experience. Additionally, AWS Marketplace also supports AWS features such as Amazon EBS-backed AMIs, Reserved Instances, and Spot Instances.

For information about how to sell your AMI on AWS Marketplace, see Selling on AWS Marketplace.

Finding a Paid AMI

There are several ways that you can find AMIs that are available for you to purchase. For example, you can use AWS Marketplace, the Amazon EC2 console, or the command line. Alternatively, a developer might let you know about a paid AMI themselves.

Finding a Paid AMI Using the Console

To find a paid AMI using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose AMIs.
3. Choose Public images for the first filter.
4. In the Search bar, choose Owner, then AWS Marketplace.
5. If you know the product code, choose Product Code, then type the product code.

Finding a Paid AMI Using AWS Marketplace

To find a paid AMI using AWS Marketplace

1. Open AWS Marketplace.
2. Enter the name of the operating system in the search box, and click Go.
3. To scope the results further, use one of the categories or filters.
4. Each product is labeled with its product type: either AMI or Software as a Service.

Finding a Paid AMI Using the Tools for Windows PowerShell

You can find a paid AMI using the following Get-EC2Image command.

```
PS C:\> Get-EC2Image -Owner aws-marketplace
```

The output for a paid AMI includes the product code.
If you know the product code, you can filter the results by product code. This example returns the most recent AMI with the specified product code.

```powershell
PS C:\> (Get-EC2Image -Owner aws-marketplace -Filter @{"Name"="product-code";"Value"="product_code"} | sort CreationDate -Descending | Select-Object -First 1).ImageId
```

### Finding a Paid AMI Using the AWS CLI

You can find a paid AMI using the following `describe-images` command (AWS CLI).

```bash
aws ec2 describe-images --owners aws-marketplace
```

This command returns numerous details that describe each AMI, including the product code for a paid AMI. The output from `describe-images` includes an entry for the product code like the following:

```json
"ProductCodes": [
  {
    "ProductCodeId": "product_code",
    "ProductCodeType": "marketplace"
  }
],
```

If you know the product code, you can filter the results by product code. This example returns the most recent AMI with the specified product code.

```bash
aws ec2 describe-images --owners aws-marketplace --filters "Name=product-code,Values=product_code" --query "sort_by(Images, &CreationDate)[-1].[ImageId]"
```

### Purchasing a Paid AMI

You must sign up for (purchase) a paid AMI before you can launch an instance using the AMI.

Typically a seller of a paid AMI presents you with information about the AMI, including its price and a link where you can buy it. When you click the link, you're first asked to log into AWS, and then you can purchase the AMI.

#### Purchasing a Paid AMI Using the Console

You can purchase a paid AMI by using the Amazon EC2 launch wizard. For more information, see Launching an AWS Marketplace Instance (p. 350).

#### Subscribing to a Product Using AWS Marketplace

To use the AWS Marketplace, you must have an AWS account. To launch instances from AWS Marketplace products, you must be signed up to use the Amazon EC2 service, and you must be subscribed to the product from which to launch the instance. There are two ways to subscribe to products in the AWS Marketplace:

- **AWS Marketplace website**: You can launch preconfigured software quickly with the 1-Click deployment feature.
Getting the Product Code for Your Instance

You can retrieve the AWS Marketplace product code for your instance using its instance metadata. For more information about retrieving metadata, see Instance Metadata and User Data (p. 478).

To retrieve a product code, use the following command:

```
```

If the instance has a product code, Amazon EC2 returns it.

Using Paid Support

Amazon EC2 also enables developers to offer support for software (or derived AMIs). Developers can create support products that you can sign up to use. During sign-up for the support product, the developer gives you a product code, which you must then associate with your own AMI. This enables the developer to confirm that your instance is eligible for support. It also ensures that when you run instances of the product, you are charged according to the terms for the product specified by the developer.

**Important**

You can't use a support product with Reserved Instances. You always pay the price that's specified by the seller of the support product.

To associate a product code with your AMI, use one of the following commands, where `ami_id` is the ID of the AMI and `product_code` is the product code:

- modify-image-attribute (AWS CLI)

  ```
  aws ec2 modify-image-attribute --image-id ami_id --product-codes "product_code"
  ```

- Edit-EC2ImageAttribute (AWS Tools for Windows PowerShell)

  ```
  PS C:\> Edit-EC2ImageAttribute -ImageId ami_id -ProductCode product_code
  ```

After you set the product code attribute, it cannot be changed or removed.

Bills for Paid and Supported AMIs

At the end of each month, you receive an email with the amount your credit card has been charged for using any paid or supported AMIs during the month. This bill is separate from your regular Amazon EC2 bill. For more information, see Paying For AWS Marketplace Products.

Managing Your AWS Marketplace Subscriptions

On the AWS Marketplace website, you can check your subscription details, view the vendor's usage instructions, manage your subscriptions, and more.

**To check your subscription details**

1. Log in to the AWS Marketplace.
Creating a Custom Windows AMI

To create a Windows AMI, you launch an instance from an existing Windows AMI, customize the instance, and create a new AMI from the instance.

To create a custom Linux AMI, use the procedure for the type of volume for the instance. For more information, see Creating an Amazon EBS-Backed Linux AMI or Creating an Instance Store-Backed Linux AMI in the Amazon EC2 User Guide for Linux Instances.

Overview of Creating an AMI

First, launch an instance from an AMI that's similar to the AMI that you'd like to create. You can connect to your instance and customize it. When the instance is set up the way you want it, ensure data integrity by stopping the instance before you create an AMI and then create the image. We automatically register the AMI for you.

During the AMI-creation process, Amazon EC2 creates snapshots of your instance's root volume and any other EBS volumes attached to your instance. You're charged for the snapshots until you deregister the AMI and delete the snapshots. For more information, see Deregistering Your Windows AMI (p. 77). If any volumes attached to the instance are encrypted, the new AMI only launches successfully on instance types that support Amazon EBS encryption. For more information, see Amazon EBS Encryption (p. 865).

Depending on the size of the volumes, it can take several minutes for the AMI-creation process to complete (sometimes up to 24 hours). You may find it more efficient to create snapshots of your volumes prior to creating your AMI. This way, only small, incremental snapshots need to be created when the AMI is created, and the process completes more quickly (the total time for snapshot creation remains the same). For more information, see Creating Amazon EBS Snapshots (p. 822).

After the process completes, you have a new AMI and snapshot created from the root volume of the instance. When you launch an instance using the new AMI, we create a new EBS volume for its root volume using the snapshot.
If you add instance store volumes or Amazon EBS volumes to your instance in addition to the root device volume, the block device mapping for the new AMI contains information for these volumes, and the block device mappings for instances that you launch from the new AMI automatically contain information for these volumes. The instance store volumes specified in the block device mapping for the new instance are new and don’t contain any data from the instance store volumes of the instance you used to create the AMI. The data on EBS volumes persists. For more information, see Block Device Mapping (p. 930).

**Note**
When you create a new instance from a custom AMI, you should initialize both its root volume and any additional EBS storage before putting it into production. For more information, see Initializing Amazon EBS Volumes.

### Creating a Windows AMI from a Running Instance

You can create an AMI using the AWS Management Console or the command line. The following diagram summarizes the process for creating an AMI from a running EC2 instance. Start with an existing AMI, launch an instance, customize it, create a new AMI from it, and finally launch an instance of your new AMI. The steps in the following diagram match the steps in the procedure below. If you already have a running Windows instance, you can go directly to step 4.

**To create an AMI from an instance using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Images, AMIs**.
3. Use the **Filter** options to scope the list of AMIs to the Windows AMIs that meet your needs. For example, to view the Windows AMIs provided by AWS, choose **Public images** from the drop-down list. Choose the Search bar. Choose **Owner** from the menu and choose **Amazon images**. Choose **Source** from the menu and type one of the following, depending on the version of Windows Server that you need:
   - amazon/Windows_Server-2019
   - amazon/Windows_Server-2016
   - amazon/Windows_Server-2012
   - amazon/Windows_Server-2008

   Add any other filters that you need. When you have chosen an AMI, select its checkbox.
4. Choose **Launch**. Accept the default values as you step through the wizard. For more information, see Launching an Instance Using the Launch Instance Wizard (p. 334). When the instance is ready, connect to it. For more information, see Connecting to Your Windows Instance (p. 379).
5. You can perform any of the following actions on your instance to customize it for your needs:
   - Install software and applications
   - Copy data
   - Reduce start time by deleting temporary files, defragmenting your hard drive, and zeroing out free space
Creating a Windows AMI from a Running Instance

- Attach additional EBS volumes
- Create a new user account and add it to the Administrators group

If you are sharing your AMI, these credentials can be supplied for RDP access without disclosing your default administrator password.

- [Windows Server 2016 and later] Configure settings using EC2Launch. To generate a random password at launch time, use the adminPasswordType setting. For more information, see Configuring EC2Launch (p. 397).
- [Windows Server 2012 R2 and earlier] Configure settings using EC2Config. To generate a random password at launch time, enable the Ec2SetPassword plugin; otherwise, the current administrator password is used. For more information, see EC2Config Settings Files (p. 411).
- [Windows Server 2008 R2] If the instance uses RedHat drivers to access Xen virtualized hardware, upgrade to Citrix drivers before you create an AMI. For more information, see Upgrade Windows Server 2008 and 2008 R2 Instances (Redhat to Citrix PV Upgrade) (p. 438).

6. In the navigation pane, choose **Instances** and select your instance. Choose **Actions, Image**, and **Create Image**.

   **Tip**
   If this option is disabled, your instance isn’t an Amazon EBS-backed instance.

7. Specify a unique name for the image and an optional description (up to 255 characters).

   By default, Amazon EC2 shuts down the instance, takes snapshots of any attached volumes, creates and registers the AMI, and then reboots the instance. Choose **No reboot** if you don’t want your instance to be shut down.

   **Warning**
   If you choose **No reboot**, we can’t guarantee the file system integrity of the created image.

   (Optional) Modify the root volume, Amazon EBS volumes, and instance store volumes as needed. For example:

   - To change the size of the root volume, locate the **Root** volume in the **Type** column, and fill in the **Size** field.
   - To suppress an Amazon EBS volume specified by the block device mapping of the AMI used to launch the instance, locate the EBS volume in the list and choose **Delete**.
   - To add an Amazon EBS volume, choose **Add New Volume, Type**, and **EBS**, and fill in the fields. When you then launch an instance from your new AMI, these additional volumes are automatically attached to the instance. Empty volumes must be formatted and mounted. Volumes based on a snapshot must be mounted.
   - To suppress an instance store volume specified by the block device mapping of the AMI used to launch the instance, locate the volume in the list and choose **Delete**.
   - To add an instance store volume, choose **Add New Volume, Type**, and **Instance Store**, and select a device name from the **Device** list. When you launch an instance from your new AMI, these additional volumes are automatically initialized and mounted. These volumes don’t contain data from the instance store volumes of the running instance from which you based your AMI.

   When you are finished, choose **Create Image**.

8. While your AMI is being created, you can choose **AMIs** in the navigation pane to view its status. Initially, this is pending. After a few minutes, the status should change to **available**.

   (Optional) Choose **Snapshots** in the navigation pane to view the snapshot that was created for the new AMI. When you launch an instance from this AMI, we use this snapshot to create its root device volume.

9. Launch an instance from your new AMI. For more information, see **Launching an Instance Using the Launch Instance Wizard** (p. 334). The new running instance contains all of the customizations you...
To create an AMI from an instance using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- create-image (AWS CLI)
- New-EC2Image (AWS Tools for Windows PowerShell)

Using Encryption with EBS-Backed AMIs

AMIs that are backed by Amazon EBS snapshots can take advantage of Amazon EBS encryption. Snapshots of both data and root volumes can be encrypted and attached to an AMI. You can launch instances and copy images with full EBS encryption support included. Encryption parameters for these operations are supported in all Regions where AWS KMS is available.

EC2 instances with encrypted EBS volumes are launched from AMIs in the same way as other instances. In addition, when you launch an instance from an AMI backed by unencrypted EBS snapshots, you can encrypt some or all of the volumes during launch.

Like EBS volumes, snapshots in AMIs can be encrypted by either your default AWS Key Management Service customer master key (CMK), or to a customer managed key that you specify. You must in all cases have permission to use the selected key.

AMIs with encrypted snapshots can be shared across AWS accounts. For more information, see Shared AMIs.

Instance-Launching Scenarios

Amazon EC2 instances are launched from AMIs using the RunInstances action with parameters supplied through block device mapping, either by means of the AWS Management Console or directly using the Amazon EC2 API or CLI. For more information about block device mapping, see Block Device Mapping. For examples of controlling block device mapping from the AWS CLI, see Launch, List, and Terminate EC2 Instances.

By default, without explicit encryption parameters, a RunInstances action maintains the existing encryption state of an AMI’s source snapshots while restoring EBS volumes from them. If encryption by default is enabled, all volumes created from the AMI (whether from encrypted or unencrypted snapshots) will be encrypted. If encryption by default is not enabled, then the instance maintains the encryption state of the AMI.

You can also launch an instance and simultaneously apply a new encryption state to the resulting volumes by supplying encryption parameters. Consequently, the following behaviors are observed:

Launch with no encryption parameters

- An unencrypted snapshot is restored to an unencrypted volume, unless encryption by default is enabled, in which case all the newly created volumes will be encrypted.
- An encrypted snapshot that you own is restored to a volume that is encrypted to the same CMK.
- An encrypted snapshot that you do not own (for example, the AMI is shared with you) is restored to a volume that is encrypted by your AWS account’s default CMK.
The default behaviors can be overridden by supplying encryption parameters. The available parameters are Encrypted and KmsKeyId. Setting only the Encrypted parameter results in the following:

**Instance launch behaviors with Encrypted set, but no KmsKeyId specified**

- An unencrypted snapshot is restored to an EBS volume that is encrypted by your AWS account's default CMK.
- An encrypted snapshot that you own is restored to an EBS volume encrypted by the same CMK. (In other words, the Encrypted parameter has no effect.)
- An encrypted snapshot that you do not own (i.e., the AMI is shared with you) is restored to a volume that is encrypted by your AWS account's default CMK. (In other words, the Encrypted parameter has no effect.)

Setting both the Encrypted and KmsKeyId parameters allows you to specify a non-default CMK for an encryption operation. The following behaviors result:

**Instance with both Encrypted and KmsKeyId set**

- An unencrypted snapshot is restored to an EBS volume encrypted by the specified CMK.
- An encrypted snapshot is restored to an EBS volume encrypted not to the original CMK, but instead to the specified CMK.

Submitting a KmsKeyId without also setting the Encrypted parameter results in an error.

The following sections provide examples of launching instances from AMIs using non-default encryption parameters. In each of these scenarios, parameters supplied to the RunInstances action result in a change of encryption state during restoration of a volume from a snapshot.

**Note**
For detailed console procedures to launch an instance from an AMI, see Launch Your Instance. For documentation of the RunInstances API, see RunInstances. For documentation of the run-instances command in the AWS Command Line Interface, see run-instances.

**Encrypt a Volume during Launch**

In this example, an AMI backed by an unencrypted snapshot is used to launch an EC2 instance with an encrypted EBS volume.
The `Encrypted` parameter alone results in the volume for this instance being encrypted. Providing a `KmsKeyId` parameter is optional. If no key ID is specified, the AWS account’s default CMK is used to encrypt the volume. To encrypt the volume to a different CMK that you own, supply the `KmsKeyId` parameter.

**Re-Encrypt a Volume during Launch**

In this example, an AMI backed by an encrypted snapshot is used to launch an EC2 instance with an EBS volume encrypted by a new CMK.

If you own the AMI and supply no encryption parameters, the resulting instance has a volume encrypted by the same key as the snapshot. If the AMI is shared rather than owned by you, and you supply no encryption parameters, the volume is encrypted by your default CMK. With encryption parameters supplied as shown, the volume is encrypted by the specified CMK.

**Change Encryption State of Multiple Volumes during Launch**

In this more complex example, an AMI backed by multiple snapshots (each with its own encryption state) is used to launch an EC2 instance with a newly encrypted volume and a re-encrypted volume.
In this scenario, the `RunInstances` action is supplied with encryption parameters for each of the source snapshots. When all possible encryption parameters are specified, the resulting instance is the same regardless of whether you own the AMI.

Image-Copying Scenarios

Amazon EC2 AMIs are copied using the `CopyImage` action, either through the AWS Management Console or directly using the Amazon EC2 API or CLI.

By default, without explicit encryption parameters, a `CopyImage` action maintains the existing encryption state of an AMI’s source snapshots during copy. You can also copy an AMI and simultaneously apply a new encryption state to its associated EBS snapshots by supplying encryption parameters. Consequently, the following behaviors are observed:

**Copy with no encryption parameters**

- An unencrypted snapshot is copied to another unencrypted snapshot, unless encryption by default is enabled, in which case all the newly created snapshots will be encrypted.
- An encrypted snapshot that you own is copied to a snapshot encrypted with the same key.
- An encrypted snapshot that you do not own (that is, the AMI is shared with you) is copied to a snapshot that is encrypted by your AWS account’s default CMK.

All of these default behaviors can be overridden by supplying encryption parameters. The available parameters are `Encrypted` and `KmsKeyId`. Setting only the `Encrypted` parameter results in the following:

**Copy-image behaviors with Encrypted set, but no KmsKeyId specified**

- An unencrypted snapshot is copied to a snapshot encrypted by the AWS account’s default CMK.
- An encrypted snapshot is copied to a snapshot encrypted by the same CMK. (In other words, the `Encrypted` parameter has no effect.)
• An encrypted snapshot that you do not own (i.e., the AMI is shared with you) is copied to a volume that is encrypted by your AWS account's default CMK. (In other words, the Encrypted parameter has no effect.)

Setting both the Encrypted and KmsKeyId parameters allows you to specify a customer managed CMK for an encryption operation. The following behaviors result:

**Copy-image behaviors with both Encrypted and KmsKeyId set**

• An unencrypted snapshot is copied to a snapshot encrypted by the specified CMK.
• An encrypted snapshot is copied to a snapshot encrypted not to the original CMK, but instead to the specified CMK.

Submitting a KmsKeyId without also setting the Encrypted parameter results in an error.

The following section provides an example of copying an AMI using non-default encryption parameters, resulting in a change of encryption state.

**Note**

For detailed console procedures to copy an AMI, see [Copying an AMI](#).
For documentation of the CopyImage API, see [CopyImage](#).
For documentation of the command `copy-image` in the AWS Command Line Interface, see [copy-image](#).

### Encrypt an Unencrypted Image during Copy

In this scenario, an AMI backed by an unencrypted root snapshot is copied to an AMI with an encrypted root snapshot. The CopyImage action is invoked with two encryption parameters, including a CMK. As a result, the encryption status of the root snapshot changes, so that the target AMI is backed by a root snapshot containing the same data as the source snapshot, but encrypted using the specified key. You incur storage costs for the snapshots in both AMIs, as well as charges for any instances you launch from either AMI.

**Note**

Enabling encryption by default (p. 868) has the same effect as setting the Encrypted parameter to `true` for all snapshots in the AMI.

Setting the Encrypted parameter encrypts the single snapshot for this instance. If you do not specify the KmsKeyId parameter, the default CMK is used to encrypt the snapshot copy.
Note
You can also copy an image with multiple snapshots and configure the encryption state of each individually.

Copying an AMI

You can copy an Amazon Machine Image (AMI) within or across AWS Regions using the AWS Management Console, the AWS Command Line Interface or SDKs, or the Amazon EC2 API, all of which support the `CopyImage` action. You can copy both Amazon EBS-backed AMIs and instance-store-backed AMIs. You can copy AMIs with encrypted snapshots and also change encryption status during the copy process.

Copying a source AMI results in an identical but distinct target AMI with its own unique identifier. In the case of an Amazon EBS-backed AMI, each of its backing snapshots is, by default, copied to an identical but distinct target snapshot. (The sole exceptions are when you choose to encrypt or re-encrypt the snapshot.) You can change or deregister the source AMI with no effect on the target AMI. The reverse is also true.

There are no charges for copying an AMI. However, standard storage and data transfer rates apply.

AWS does not copy launch permissions, user-defined tags, or Amazon S3 bucket permissions from the source AMI to the new AMI. After the copy operation is complete, you can apply launch permissions, user-defined tags, and Amazon S3 bucket permissions to the new AMI.

You can't copy an AMI that was obtained from the AWS Marketplace, regardless of whether you obtained it directly or it was shared with you. Instead, launch an EC2 instance using the AWS Marketplace AMI and then create an AMI from the instance. For more information, see Creating a Custom Windows AMI (p. 64).

Permissions for Copying an Instance Store-Backed AMI

If you use an IAM user to copy an instance store-backed AMI, the user must have the following Amazon S3 permissions: `s3:CreateBucket`, `s3:GetBucketAcl`, `s3:ListAllMyBuckets`, `s3:GetObject`, `s3:PutObject`, and `s3:PutObjectAcl`.

The following example policy allows the user to copy the AMI source in the specified bucket to the specified Region.

```json
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": "s3:ListAllMyBuckets",
         "Resource": [
            "arn:aws:s3:::*"
         ]
      },
      {
         "Effect": "Allow",
         "Action": "s3:GetObject",
         "Resource": [
            "arn:aws:s3:::ami-source-bucket/*"
         ]
      }
   ]
}
```
Cross-Region Copying

Copying an AMI across geographically diverse Regions provides the following benefits:

- **Consistent global deployment:** Copying an AMI from one Region to another enables you to launch consistent instances in different Regions based on the same AMI.
- **Scalability:** You can more easily design and build global applications that meet the needs of your users, regardless of their location.
- **Performance:** You can increase performance by distributing your application, as well as locating critical components of your application in closer proximity to your users. You can also take advantage of Region-specific features, such as instance types or other AWS services.
- **High availability:** You can design and deploy applications across AWS regions, to increase availability.

The following diagram shows the relations among a source AMI and two copied AMIs in different Regions, as well as the EC2 instances launched from each. When you launch an instance from an AMI, it resides in the same Region where the AMI resides. If you make changes to the source AMI and want those changes to be reflected in the AMIs in the target Regions, you must recopy the source AMI to the target Regions.

When you first copy an instance store-backed AMI to a Region, we create an Amazon S3 bucket for the AMIs copied to that Region. All instance store-backed AMIs that you copy to that Region are stored in this bucket. The bucket names have the following format: amis-for-account-in-region-hash. For example: amis-for-123456789012-in-us-east-2-yhjmxvp6.
Prerequisite

Prior to copying an AMI, you must ensure that the contents of the source AMI are updated to support running in a different Region. For example, you should update any database connection strings or similar application configuration data to point to the appropriate resources. Otherwise, instances launched from the new AMI in the destination Region may still use the resources from the source Region, which can impact performance and cost.

Limits

- Destination Regions are limited to 50 concurrent AMI copies.

Cross-Account Copying

You can share an AMI with another AWS account. Sharing an AMI does not affect the ownership of the AMI. The owning account is charged for the storage in the Region. For more information, see Sharing an AMI with Specific AWS Accounts (p. 57).

If you copy an AMI that has been shared with your account, you are the owner of the target AMI in your account. The owner of the source AMI is charged standard Amazon EBS or Amazon S3 transfer fees, and you are charged for the storage of the target AMI in the destination Region.

Resource Permissions

To copy an AMI that was shared with you from another account, the owner of the source AMI must grant you read permissions for the storage that backs the AMI, either the associated EBS snapshot (for an Amazon EBS-backed AMI) or an associated S3 bucket (for an instance store-backed AMI). If the shared AMI has encrypted snapshots, the owner must share the key or keys with you as well.

Encryption and Copying

The following table shows encryption support for various AMI-copying scenarios. While it is possible to copy an unencrypted snapshot to yield an encrypted snapshot, you cannot copy an encrypted snapshot to yield an unencrypted one.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unencrypted-to-unencrypted</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Encrypted-to-encrypted</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Unencrypted-to-encrypted</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Encrypted-to-unencrypted</td>
<td>No</td>
</tr>
</tbody>
</table>

Note

Encrypting during the CopyImage action applies only to Amazon EBS-backed AMIs. Because an instance store-backed AMI does not rely on snapshots, you cannot use copying to change its encryption status.

By default (i.e., without specifying encryption parameters), the backing snapshot of an AMI is copied with its original encryption status. Copying an AMI backed by an unencrypted snapshot results in an identical target snapshot that is also unencrypted. If the source AMI is backed by an encrypted snapshot, copying it results in an identical target snapshot that is encrypted by the same customer master key.
(CMK). Copying an AMI backed by multiple snapshots preserves, by default, the source encryption status in each target snapshot.

If you specify encryption parameters while copying an AMI, you can encrypt or re-encrypt its backing snapshots. The following example shows a non-default case that supplies encryption parameters to the `CopyImage` action in order to change the target AMI's encryption state.

**Copy an unencrypted source AMI to an encrypted target AMI**

In this scenario, an AMI backed by an unencrypted root snapshot is copied to an AMI with an encrypted root snapshot. The `CopyImage` action is invoked with two encryption parameters, including a CMK. As a result, the encryption status of the root snapshot changes, so that the target AMI is backed by a root snapshot containing the same data as the source snapshot, but encrypted using the specified key. You incur storage costs for the snapshots in both AMIs, as well as charges for any instances you launch from either AMI.

**Note**

Enabling encryption by default (p. 868) has the same effect as setting the `Encrypted` parameter to `true` for all snapshots in the AMI.

![Diagram of CopyImage process]

Setting the `Encrypted` parameter encrypts the single snapshot for this instance. If you do not specify the `KmsKeyId` parameter, the default CMK is used to encrypt the snapshot copy.

For more information about copying AMIs with encrypted snapshots, see Using Encryption with EBS-Backed AMIs (p. 67).

## Copying an AMI

You can copy an AMI as follows.

**Prerequisite**

Create or obtain an AMI backed by an Amazon EBS snapshot. Note that you can use the Amazon EC2 console to search a wide variety of AMIs provided by AWS. For more information, see Creating a Custom Windows AMI (p. 64) and Finding a Windows AMI (p. 51).

**To copy an AMI using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the console navigation bar, select the Region that contains the AMI. In the navigation pane, choose Images, AMIs to display the list of AMIs available to you in the Region.

3. Select the AMI to copy and choose Actions, Copy AMI.

4. In the Copy AMI dialog box, specify the following information and then choose Copy AMI:

   - **Destination region**: The Region in which to copy the AMI.
   - **Name**: A name for the new AMI. You can include operating system information in the name, as we do not provide this information when displaying details about the AMI.
   - **Description**: By default, the description includes information about the source AMI so that you can distinguish a copy from its original. You can change this description as needed.
   - **Encryption**: Select this field to encrypt the target snapshots, or to re-encrypt them using a different key. If you have enabled encryption by default, the Encryption option is set and cannot be unset from the AMI console.
   - **Master Key**: The KMS key to used to encrypt the target snapshots.

5. We display a confirmation page to let you know that the copy operation has been initiated and to provide you with the ID of the new AMI.

   To check on the progress of the copy operation immediately, follow the provided link. To check on the progress later, choose Done, and then when you are ready, use the navigation bar to switch to the target region (if applicable) and locate your AMI in the list of AMIs.

   The initial status of the target AMI is pending and the operation is complete when the status is available.

**To copy an AMI using the AWS CLI**

You can copy an AMI using the copy-image command. You must specify both the source and destination Regions. You specify the source Region using the --source-region parameter. You can specify the destination Region using either the --region parameter or an environment variable. For more information, see Configuring the AWS Command Line Interface.

When you encrypt a target snapshot during copying, you must specify these additional parameters: --encrypted and --kms-key-id.

**To copy an AMI using the Tools for Windows PowerShell**

You can copy an AMI using the Copy-EC2Image command. You must specify both the source and destination Regions. You specify the source Region using the -SourceRegion parameter. You can specify the destination Region using either the -Region parameter or the Set-AWSDefaultRegion command. For more information, see Specifying AWS Regions.

When you encrypt a target snapshot during copying, you must specify these additional parameters: -Encrypted and -KmsKeyId.

**Stopping a Pending AMI Copy Operation**

You can stop a pending AMI copy as follows.

**To stop an AMI copy operation using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the destination Region from the Region selector.
3. In the navigation pane, choose AMIs.
4. Select the AMI to stop copying and choose Actions, Deregister.
5. When asked for confirmation, choose **Continue**.

**To stop an AMI copy operation using the command line**

You can use one of the following commands. For more information about these command line interfaces, see [Accessing Amazon EC2 (p. 3)](#).

- **deregister-image** (AWS CLI)
- **Unregister-EC2Image** (AWS Tools for Windows PowerShell)

**Deregistering Your Windows AMI**

You can deregister a Windows AMI when you have finished using it. After you deregister an AMI, you can’t use it to launch new instances.

When you deregister an AMI, it doesn’t affect any instances that you’ve already launched from the AMI or any snapshots created for the EBS root volume during the AMI creation process. You’ll continue to incur usage costs for these instances and storage costs for the snapshot. Therefore, you should terminate any instances that you finished with and delete any snapshots that you are finished with.

The following diagram illustrates the process for cleaning up your Windows AMI.

![Diagram](image)

**To clean up your Windows AMI**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **AMIs**. Select the AMI and take note of its ID — this can help you find the correct snapshot in the next step. Choose **Actions**, and then **Deregister**. When prompted for confirmation, choose **Continue**.

   **Note**
   It can take a few minutes before the console removes the AMI from the list. Choose **Refresh** to refresh the status.

3. In the navigation pane, choose **Snapshots**, and select the snapshot (look for the AMI ID in the **Description** column). Choose **Actions**, and then choose **Delete Snapshot**. When prompted for confirmation, choose **Yes, Delete**.
4. (Optional) If you are finished with an instance that you launched from the AMI, terminate it. In the navigation pane, choose **Instances**. Select the instance, choose **Actions**, then **Instance State**, and then **Terminate**. When prompted for confirmation, choose **Yes, Terminate**.

### Managed AWS Windows AMIs

AWS provides managed Amazon Machine Images (AMIs) that include various versions and configurations of Windows Server. In general, the AWS Windows AMIs are configured with the default settings used by the Microsoft installation media. However, there are customizations. For example, the AWS Windows AMIs come with the following software and drivers:

- EC2Config service (through Windows Server 2012 R2)
- EC2Launch (Windows Server 2016 and later)
- AWS Systems Manager
- AWS CloudFormation
- AWS Tools for Windows PowerShell
- Network drivers (SRIOV, ENA, Citrix PV)
- Storage drivers (NVMe, AWS PV, Citrix PV)
- Graphics drivers (NVidia GPU, Elastic GPU)
- Spot Instance hibernation

For information about other customizations, see [Configuration Changes for AWS Windows AMIs](#) (p. 80).

**Contents**

- Updating Your Windows Instance (p. 79)
- Upgrading or Migrating to a Newer Version of Windows Server (p. 79)
- Subscribing to Windows AMI Notifications (p. 79)
- Configuration Changes for AWS Windows AMIs (p. 80)
- Details About AWS Windows AMI Versions (p. 82)
  - What to Expect in an Official AWS Windows AMI (p. 82)
  - How AWS Decides Which Windows AMIs to Offer (p. 83)
  - Patches, Security Updates, and AMI IDs (p. 83)
  - Semiannual Channel Releases (p. 84)
  - Monthly AMI Updates for 2019 (to date) (p. 84)
  - Monthly AMI Updates for 2018 (p. 88)
  - Monthly AMI Updates for 2017 (p. 94)
  - Monthly AMI Updates for 2016 (p. 98)
  - Monthly AMI Updates for 2015 (p. 101)
  - Monthly AMI Updates for 2014 (p. 104)
  - Monthly AMI Updates for 2013 (p. 105)
  - Monthly AMI Updates for 2012 (p. 107)
  - Monthly AMI Updates for 2011 and earlier (p. 109)
- Changes in Windows Server 2016 and Later AMIs (p. 109)
- Docker Container Conflict on Windows Server 2016 Instances (p. 110)
- Issue with the Hibernate Agent (2018.03.16 AMIs) (p. 110)
Updating Your Windows Instance

After you launch a Windows instance, you are responsible for installing updates on it. You can manually install only the updates that interest you, or you can start from a current AWS Windows AMI and build a new Windows instance. For information about finding the current AWS Windows AMIs, see Finding a Windows AMI (p. 51).

Note
Instances should be stateless when updating. For more information, see Managing Your AWS Infrastructure at Scale.

For Windows instances, you can install updates to the following services or applications:

- Microsoft Windows Server
- Microsoft SQL Server
- Windows PowerShell
- EC2Launch (p. 396)
- EC2Config service (p. 406)
- AWS Systems Manager SSM Agent
- ENA (p. 702)
- NVMe Drivers (p. 446)
- PV Drivers (p. 436)
- AWS Tools for Windows PowerShell
- AWS CloudFormation helper scripts

You can reboot a Windows instance after installing updates. For more information, see Reboot Your Instance (p. 385).

Upgrading or Migrating to a Newer Version of Windows Server

For information about how to upgrade or migrate a Windows instance to a newer version of Windows Server, see Upgrading an Amazon EC2 Windows Instance to a Newer Version of Windows Server (p. 495).

Subscribing to Windows AMI Notifications

To be notified when new AMIs are released or when previously released AMIs are made private, subscribe for notifications using Amazon SNS.

To subscribe to Windows AMI notifications
2. In the navigation bar, change the region to **US East (N. Virginia)**, if necessary. You must use this region because the SNS notifications that you are subscribing to were created in this region.
3. In the navigation pane, choose **Subscriptions**.
4. Choose **Create subscription**.
5. For the **Create subscription** dialog box, do the following:
   a. For **Topic ARN**, copy and paste one of the following Amazon Resource Names (ARNs):
      - **arn:aws:sns:us-east-1:801119661308:ec2-windows-ami-update**
      - **arn:aws:sns:us-east-1:801119661308:ec2-windows-ami-private**
      For **AWS GovCloud (US)**:
   b. For **Protocol**, choose **Email**.
   c. For **Endpoint**, type an email address that you can use to receive the notifications.
   d. Choose **Create subscription**.
6. You'll receive a confirmation email with the subject line **AWS Notification - Subscription Confirmation**. Open the email and choose **Confirm subscription** to complete your subscription.

Whenever Windows AMIs are released, we send notifications to the subscribers of the `ec2-windows-ami-update` topic. Whenever released Windows AMIs are made private, we send notifications to the subscribers of the `ec2-windows-ami-private` topic. If you no longer want to receive these notifications, use the following procedure to unsubscribe.

**To unsubscribe from Windows AMI notifications**
2. In the navigation bar, change the region to **US East (N. Virginia)**, if necessary. You must use this region because the SNS notifications were created in this region.
3. In the navigation pane, choose **Subscriptions**.
4. Select the subscriptions and then choose **Actions, Delete subscriptions** When prompted for confirmation, choose **Delete**.

**Configuration Changes for AWS Windows AMIs**

The following changes are applied to each AWS Windows AMI.

**Clean and Prepare**

<table>
<thead>
<tr>
<th>Change</th>
<th>Applies to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check for pending file renames or reboots, and reboot as needed</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Delete .dmp files</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Delete logs (event logs, Systems Manager, EC2Config)</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Delete temporary folders and files for sysprep</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Clear recent history (Start menu, Windows Explorer, and so on)</td>
<td>Windows Server 2012 R2 and earlier</td>
</tr>
<tr>
<td>Perform virus scan</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Change</td>
<td>Applies to</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Pre-compile queued .NET assemblies (before sysprep)</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Run Windows maintenance tools</td>
<td>Windows Server 2012 R2 and later</td>
</tr>
<tr>
<td>Restore default values for Internet Explorer</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Restore default values for EC2Config</td>
<td>Windows Server 2012 R2 and earlier</td>
</tr>
<tr>
<td>Set EC2Launch to run at the next launch</td>
<td>Windows Server 2016 and later</td>
</tr>
<tr>
<td>Reset the Windows wallpaper</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Run sysprep</td>
<td>All AMIs</td>
</tr>
</tbody>
</table>

### Install and Configure

<table>
<thead>
<tr>
<th>Change</th>
<th>Applies to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add links to the Amazon EC2 Windows Guide</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Attach instance storage volumes to extended mount points</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Install the current AWS Tools for Windows PowerShell</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Install the current AWS CloudFormation helper scripts</td>
<td>All AMIs</td>
</tr>
<tr>
<td>Install the current EC2Config and SSM Agent</td>
<td>Windows Server 2012 R2 and earlier</td>
</tr>
<tr>
<td>Install the current EC2Launch and SSM Agent</td>
<td>Windows Server 2016 and later</td>
</tr>
<tr>
<td>Install the current AWS PV, ENA, and NVMe drivers</td>
<td>Windows Server 2008 R2 and later</td>
</tr>
<tr>
<td>Install the current SRIOV drivers</td>
<td>Windows Server 2012 R2 and later</td>
</tr>
<tr>
<td>Install the current Citrix PV driver</td>
<td>Windows Server 2008 SP2 and earlier</td>
</tr>
<tr>
<td>Install the current EC2WinUtil driver</td>
<td>Windows Server 2008 R2 and later</td>
</tr>
<tr>
<td>Install PowerShell 2.0 and 3.0</td>
<td>Windows Server 2008 SP2 and R2</td>
</tr>
<tr>
<td>If Microsoft SQL Server is installed:</td>
<td>All AMIs</td>
</tr>
<tr>
<td>• Install service packs</td>
<td></td>
</tr>
<tr>
<td>• Configure to start automatically</td>
<td></td>
</tr>
<tr>
<td>• Add BUILTIN\Administrators to the SysAdmin role</td>
<td></td>
</tr>
<tr>
<td>• Open TCP port 1433 and UDP port 1434</td>
<td></td>
</tr>
<tr>
<td>Apply the following hotfixes:</td>
<td>Windows Server 2008 SP2 and R2</td>
</tr>
<tr>
<td>• MS15-011</td>
<td></td>
</tr>
<tr>
<td>Change</td>
<td>Applies to</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
</tbody>
</table>
| • KB2582281  
• KB2634328  
• KB2800213  
• KB2922223  
• KB2394911  
• KB2780879 | Windows Server 2012 R2 and earlier |
| Allow ICMP traffic through the firewall | Windows Server 2012 R2 and earlier |
| Enable file and printer sharing | Windows Server 2012 R2 and earlier |
| Disable RunOnce for Internet Explorer | All AMIs |
| Enable remote PowerShell | All AMIs |
| Configure a paging file on the system volume as follows:  
• Windows Server 2019 - Managed by the system  
• Windows Server 2016 -Managed by the system  
• Windows Server 2012 R2 - Initial size and max size are 8 GB  
• Windows Server 2012 and earlier - Initial size is 512 MB, max size is 8 GB | All AMIs |
| Configure an additional system managed paging file on Z:, if available | Windows Server 2012 R2 and earlier |
| Disable hibernation and delete the hibernation file | All AMIs |
| Set the performance options for best performance | All AMIs |
| Set the power setting to high performance | All AMIs |
| Disable the screen saver password | All AMIs |
| Set the RealTimeIsUniversal registry key | All AMIs |
| Set the timezone to UTC | All AMIs |
| Disable Windows updates and notifications | All AMIs |
| Run Windows Update and reboot until there are no pending updates | All AMIs |
| Set the display in all power schemes to never turn off | All AMIs |
| Set the PowerShell execution policy to "Unrestricted" | All AMIs |

Details About AWS Windows AMI Versions

What to Expect in an Official AWS Windows AMI

AWS provides AMIs with a variety of configurations for all supported Windows Operating System versions. For each of these images, AWS:
• Installs all Microsoft recommended Windows security patches. We release images shortly after the monthly Microsoft patches are made available.

• Installs the latest drivers for AWS hardware, including network and disk drivers, EC2WinUtil for troubleshooting, as well as GPU drivers in selected AMIs.

• Includes AWS helper software, like EC2 Config (p. 405) for Server 2012 R2 and earlier, or EC2 Launch (p. 395) for Server 2016 and later.

• Configures Windows Time to use the AWS Time Service (p. 462).

• Makes changes in all power schemes to set the display to never turn off.

• Performs minor bug fixes—generally one-line registry changes to enable or disable features that we have found to improve performance on AWS.

Other than the adjustments listed above, we keep our AMIs as close as possible to the default install. This means we default to the “stock” PowerShell or .NET framework versions, don’t install Windows Features, and generally don’t change the AMI.

How AWS Decides Which Windows AMIs to Offer

Each AMI is extensively tested prior to release to the general public. We periodically streamline our AMI offerings to simplify customer choice and to reduce costs.

• New AMI offerings are created for new OS releases. You can count on AWS releasing “Base,” “Core/Container,” and “SQL Express/Standard/Web/Enterprise” offerings in English and other widely used languages. The primary difference between Base and Core offerings is that Base offerings have a desktop/GUI whereas Core offerings are PowerShell command line only. For more information about Windows Server Core, see https://docs.microsoft.com/en-us/windows-server/administration/server-core/what-is-server-core.

• New AMI offerings are created to support new platforms—for example, the Deep Learning and “NVidia” AMIs were created to support customers using our GPU-based instance types (P2 and P3, G2 and G3, etc.).

• Less popular AMIs are sometimes removed. If we see a particular AMI is launched only a few times in its entire lifespan, we will remove it in favor of more widely used options.

If there is an AMI variant that you would like to see, let us know by filing a ticket with Cloud Support, or by providing feedback through one of our established channels.

Patches, Security Updates, and AMI IDs

AWS provides updated, fully-patched Windows AMIs within five business days of Microsoft’s patch Tuesday (the second Tuesday of each month). The new AMIs are available immediately through the Images page in the Amazon EC2 console. The new AMIs are available in the AWS Marketplace and the Quick Start tab of the launch instance wizard within a few days of their release.

Note
Instances launched from the latest Windows Server 2019 AMIs may show a Windows Update dialog message stating "Some settings are managed by your organization." This message appears as a result of changes in Windows Server 2019 and does not impact the behavior of Windows Update or your ability to manage update settings.
To remove this warning, see "Some settings are managed by your organization".

To ensure that customers have the latest security updates by default, AWS keeps Windows AMIs available for three months. After releasing new Windows AMIs, AWS makes the Windows AMIs that are older than three months private within 10 days. After an AMI has been made private, if you look at an instance launched from that AMI in the console, the AMI ID field states, "Cannot load detail for ami-xxxxx. You may not be permitted to view it." You can still retrieve the AMI ID using the AWS CLI or an AWS SDK.
The Windows AMIs in each release have new AMI IDs. Therefore, we recommend that you write scripts that locate the latest AWS Windows AMIs by their names, rather than by their IDs. For more information, see the following examples:

- Get-EC2ImageByName (AWS Tools for Windows PowerShell)
- Query for the Latest Windows AMI Using Systems Manager Parameter Store
- Walkthrough: Looking Up Amazon Machine Image IDs (AWS Lambda, AWS CloudFormation)

Semiannual Channel Releases

AWS provides Windows Server semiannual channel releases that combine the scale, performance, and elasticity of AWS with the new capabilities in the Semiannual channel release versions of Windows Server.

The following tables summarize the changes to each release of the AWS Windows AMIs. Note that some changes apply to all AWS Windows AMIs while others apply to only a subset of these AMIs.

Contents

- Monthly AMI Updates for 2019 (to date) (p. 84)
- Monthly AMI Updates for 2018 (p. 88)
- Monthly AMI Updates for 2017 (p. 94)
- Monthly AMI Updates for 2016 (p. 98)
- Monthly AMI Updates for 2015 (p. 101)
- Monthly AMI Updates for 2014 (p. 104)
- Monthly AMI Updates for 2013 (p. 105)
- Monthly AMI Updates for 2012 (p. 107)
- Monthly AMI Updates for 2011 and earlier (p. 109)

For more information about components included in these AMIs, see the following:

- EC2Config Version History (p. 417)
- EC2Launch Version History (p. 403)
- Systems Manager SSM Agent Release Notes
- Amazon ENA Driver Versions (p. 703)
- AWS PV Driver Package History (p. 432)

Monthly AMI Updates for 2019 (to date)

For more information about Microsoft updates, see Description of Software Update Services and Windows Server Update Services changes in content for 2019.

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019.09.12</td>
<td>New Windows AMI</td>
</tr>
<tr>
<td></td>
<td>• amzn2-ami-hvm-2.0.20190618-x86_64-gp2-mono</td>
</tr>
<tr>
<td></td>
<td>.NET Core 2.2, Mono 5.18, and PowerShell 6.2 pre-installed to run your .NET applications on Amazon Linux 2 with Long Term Support (LTS)</td>
</tr>
</tbody>
</table>
## Details About AWS Windows AMI Versions

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2019.09.11</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
</tbody>
</table>
| | • AWS PV driver version 8.3.2  
| | • AWS NVMe driver version 1.3.2  
| | • AWS Tools for Windows PowerShell version 3.15.826  
| | • NLA enabled on all OS 2012 RTM to 2019 AMIs  
| | • Intel 82599 VF driver reverted to version 2.0.210.0 (Server 2016) or version 2.1.138.0 (Server 2019) due to customer reported issues. Engagement with Intel concerning these issues ongoing.  
| | • Windows security updates current to September 10th, 2019  
| | Previous versions of AMIs have been marked private.  
| **New Windows AMIs** | New STIG-compliant AMIs available:  
| | • Windows_Server-2012-R2-English-STIG-Full  
| | • Windows_Server-2012-R2-English-STIG-Core  
| | • Windows_Server-2016-English-STIG-Full  
| | • Windows_Server-2016-English-STIG-Core  
| | • Windows_Server-2019-English-STIG-Full  
| | • Windows_Server-2019-English-STIG-Core  
| | **Note**  
| | NLA is now enabled on all 2012 RTM, 2012 R2, and 2016 AMIs to increase default RDP security posture. NLA remains enabled on 2019 AMIs. |
| **2019.08.16** | **All AMIs** |
| | • EC2Config version 4.9.3519  
| | • SSM Agent version 2.3.634.0  
| | • AWS Tools for PowerShell version 3.15.802  
| | • Windows Defender platform update blocked via registry due to SFC failures introduced by update. Update will be re-enabled when new patch is available.  
| | **Note**  
<p>| | Starting in September, NLA will be enabled on all 2012 RTM, 2012 R2, and 2016 AMIs to increase default RDP security posture. |</p>
<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019.07.19</td>
<td><strong>New Windows AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-ECS_Optimized-2019.07.19</td>
</tr>
<tr>
<td>2019.07.12</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to July 9th, 2019</td>
</tr>
<tr>
<td>2019.06.12</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to June 11th, 2019</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.756</td>
</tr>
<tr>
<td></td>
<td>• AWS PV driver version 8.2.7</td>
</tr>
<tr>
<td></td>
<td>• AWS NVMe driver version 1.3.1</td>
</tr>
<tr>
<td></td>
<td>• The following &quot;P3&quot; AMIs will be renamed as &quot;Tesla&quot; AMIs. These AMIs</td>
</tr>
<tr>
<td></td>
<td>will support all GPU-backed AWS instances using the Tesla driver. P3</td>
</tr>
<tr>
<td></td>
<td>AMIs will no longer be updated after this release and will be removed</td>
</tr>
<tr>
<td></td>
<td>as part of our regular cycle.</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2012-R2_RTM-English-P3-2019.06.12 replaced with</td>
</tr>
<tr>
<td></td>
<td>Windows_Server-2012-R2_RTM-English-Tesla-2019.06.12</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-P3-2019.06.12 replaced with</td>
</tr>
<tr>
<td></td>
<td>Windows_Server-2016-English-Tesla-2019.06.12</td>
</tr>
<tr>
<td></td>
<td><strong>New Windows AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>Previous versions of AMIs have been marked private.</td>
</tr>
<tr>
<td>2019.05.21</td>
<td><strong>Windows Server, version 1903</strong></td>
</tr>
<tr>
<td></td>
<td>• AMIs are now available</td>
</tr>
<tr>
<td>2019.05.15</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to May 14th, 2019</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.3429</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent version 2.3.542.0</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.735</td>
</tr>
<tr>
<td>2019.04.26</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Fixed AMIs for Windows Server 2019 with SQL to address edge cases</td>
</tr>
<tr>
<td></td>
<td>where the first launch of an instance may result in Instance</td>
</tr>
<tr>
<td></td>
<td>Impairment and Windows displays the message &quot;Please wait for the</td>
</tr>
<tr>
<td></td>
<td>User Profile Service&quot;</td>
</tr>
<tr>
<td>2019.04.21</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver rollback to version 8.2.6 from version 8.3.0</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>2019.04.10</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to April 9, 2019</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.715</td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver version 8.3.0</td>
</tr>
<tr>
<td></td>
<td>• EC2Launch version 1.3.2001360</td>
</tr>
<tr>
<td></td>
<td><strong>New Windows AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-SQL_2012_SP4_Standard-2019.04.10</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-SQL_2014_SP3_Standard-2019.04.10</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-SQL_2014_SP3_Enterprise-2019.04.10</td>
</tr>
<tr>
<td>2019.03.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to March 12, 2019</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.693</td>
</tr>
<tr>
<td></td>
<td>• EC2Launch version 1.3.2001220</td>
</tr>
<tr>
<td></td>
<td>• NVIDIA Tesla driver version 412.29 for Deep Learning and P3 AMIs (<a href="https://nvidia.custhelp.com/app/answers/detail/a_id/4772">https://nvidia.custhelp.com/app/answers/detail/a_id/4772</a>)</td>
</tr>
<tr>
<td></td>
<td>Previous versions of AMIs have been marked private</td>
</tr>
<tr>
<td>2019.02.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to February 12, 2019</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent version 2.3.444.0</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.666</td>
</tr>
<tr>
<td></td>
<td>• EC2Launch version 1.3.2001040</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.3289</td>
</tr>
<tr>
<td></td>
<td>• AWS PV driver 8.2.6</td>
</tr>
<tr>
<td></td>
<td>• EBS NVMe tool</td>
</tr>
<tr>
<td></td>
<td>SQL 2014 with Service Pack 2 and SQL 2016 with Service Pack 1 will no longer be updated after this release.</td>
</tr>
<tr>
<td>2019.02.09</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Windows AMIs have been updated. New AMIs can be found with the following date versions:</td>
</tr>
<tr>
<td></td>
<td>November &quot;2018.11.29&quot;</td>
</tr>
<tr>
<td></td>
<td>December &quot;2018.12.13&quot;</td>
</tr>
<tr>
<td></td>
<td>January &quot;2019.02.09&quot;</td>
</tr>
<tr>
<td></td>
<td>Previous versions of AMIs have been marked private</td>
</tr>
</tbody>
</table>
## Details About AWS Windows AMI Versions

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2019.01.10</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to January 10, 2019</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent version 2.3.344.0</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.647</td>
</tr>
<tr>
<td></td>
<td>• EC2Launch version 1.3.2000930</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.3160</td>
</tr>
</tbody>
</table>

**All AMIs with SQL Server**

• Latest cumulative updates

### Monthly AMI Updates for 2018

For more information about Microsoft updates, see Description of Software Update Services and Windows Server Update Services changes in content for 2018.

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2018.12.12</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to December 12, 2018</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent version 2.3.274.0</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.629</td>
</tr>
<tr>
<td></td>
<td>• EC2Launch version 1.3.2000760</td>
</tr>
</tbody>
</table>

**New Windows AMIs**

• Windows_Server-2012-R2_RTM-English-64Bit-SQL_2014_SP3_Express-2018.12.12
<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Windows_Server-2016-Spanish-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Portuguese_Portugal-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Italian-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Swedish-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-English-Core-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Hungarian-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Polish-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Turkish-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Russian-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Czech-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-French-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>• Windows_Server-2016-Portuguese_Brazil-Full-Base-2018.12.12</td>
<td></td>
</tr>
<tr>
<td>Release Date</td>
<td>Changes</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>2018.11.28</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• SSM Agent version 3.235.0</td>
</tr>
<tr>
<td></td>
<td>• Changes in all power schemes to set the display to never turn off</td>
</tr>
<tr>
<td>2018.11.20</td>
<td><strong>Windows_Server-2016-English-Deep-Learning</strong></td>
</tr>
<tr>
<td></td>
<td>• TensorFlow version 1.12</td>
</tr>
<tr>
<td></td>
<td>• MXNet version 1.3</td>
</tr>
<tr>
<td></td>
<td>• NVIDIA version 392.05</td>
</tr>
<tr>
<td>2018.11.19</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to November 19, 2018</td>
</tr>
<tr>
<td></td>
<td>• AWS SDK version 3.15.602.0</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent version 2.3.193.0</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.3067</td>
</tr>
<tr>
<td></td>
<td>• Intel Chipset INF configurations to support new instance types</td>
</tr>
</tbody>
</table>

**Windows Server, version 1809**

• AMIs are now available.
<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018.10.14</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to October 9, 2018</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell version 3.3.365.0</td>
</tr>
<tr>
<td></td>
<td>• CloudFormation version 1.4.31</td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver version 8.2.4</td>
</tr>
<tr>
<td></td>
<td>• AWS PCI Serial Driver version 1.0.0.0 (support for Windows 2008R2 and 2012 on Bare Metal instances)</td>
</tr>
<tr>
<td></td>
<td>• ENA Driver version 1.5.0</td>
</tr>
</tbody>
</table>

**Microsoft Windows Server 2016 Datacenter and Standard Editions for Nano Server**

Microsoft ended mainstream support for Windows Server 2016 Datacenter and Standard Editions for Nano Server installation options as of April 10, 2018 (see the Microsoft Support Lifecycle page for more details).

Additional information about Nano Server lifecycle, including details on launching Nano Server as a container, can be found here: https://docs.microsoft.com/en-us/windows-server/get-started/nano-in-semi-annual-channel.

<table>
<thead>
<tr>
<th>2018.09.15</th>
<th><strong>All AMIs</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Microsoft security updates current to September 12, 2018</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell version 3.3.343</td>
</tr>
<tr>
<td></td>
<td>• EC2Launch version 1.3.2000430</td>
</tr>
<tr>
<td></td>
<td>• AWS NVMe Driver version 1.3.0</td>
</tr>
<tr>
<td></td>
<td>• EC2 WinUtil Driver version 2.0.0</td>
</tr>
</tbody>
</table>

**Microsoft Windows Server 2016 Base Nano**

Access to all public versions of Windows_Server-2016-English-Nano-Base will be removed in September 2018. Additional information about Nano Server lifecycle, including details on launching Nano Server as a Container, can be found here: https://docs.microsoft.com/en-us/windows-server/get-started/nano-in-semi-annual-channel.
## Details About AWS Windows AMI Versions

### 2018.08.15

**All AMIs**
- Microsoft security updates current to August 14, 2018
- AWS Tools for Windows PowerShell version 3.3.335
- AMIs now default to use Amazon's NTP service at IP 169.254.169.123 for time synchronization. For more information, see Default NTP Settings for Amazon Windows AMIs (p. 462).

**Microsoft Windows Server 2016 Base Nano**
Access to all public versions of Windows_Server-2016-English-Nano-Base will be removed in September 2018. Additional information about Nano Server lifecycle, including details on launching Nano Server as a Container, can be found here: [https://docs.microsoft.com/en-us/windows-server/get-started/nano-in-semi-annual-channel](https://docs.microsoft.com/en-us/windows-server/get-started/nano-in-semi-annual-channel).

### 2018.07.11

**All AMIs**
- Microsoft security updates current to July 10, 2018
- EC2Config version 4.9.2756
- SSM Agent 2.2.800.0

### 2018.06.22

**Windows Server 2008 R2**
- Resolves an issue with the 2018.06.13 AMIs when changing an instance from a previous generation to a current generation (for example, M4 to M5).

### 2018.06.13

**All AMIs**
- Microsoft security updates current to June 12, 2018
- EC2Config version 4.9.2688
- SSM Agent 2.2.619.0
- AWS Tools for Windows PowerShell 3.3.283.0
- AWS NVMe driver 1.2.0
- AWS PV driver 8.2.3

### 2018.05.09

**All AMIs**
- Microsoft security updates current to May 9, 2018
- EC2Config version 4.9.2644
- SSM Agent 2.2.493.0
- AWS Tools for Windows PowerShell 3.3.270.0

**Windows Server, version 1709 and Windows Server, version 1803**
- AMIs are now available. For more information, see Windows Server version 1709 and 1803 AMIs for Amazon EC2.
<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018.04.11</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to April 10, 2018</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2586</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.2.392.0</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.256.0</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation templates 1.4.30</td>
</tr>
<tr>
<td></td>
<td>• Serial INF and Intel Chipset INF configurations to support new instance types</td>
</tr>
<tr>
<td></td>
<td><strong>SQL Server 2017</strong></td>
</tr>
<tr>
<td></td>
<td>• Cumulative update 5 (CU5)</td>
</tr>
<tr>
<td></td>
<td><strong>SQL Server 2016 SP1</strong></td>
</tr>
<tr>
<td></td>
<td>• Cumulative update 8 (CU8)</td>
</tr>
<tr>
<td>2018.03.24</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to March 13, 2018</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2565</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.2.355.0</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.245.0</td>
</tr>
<tr>
<td></td>
<td>• AWS PV driver 8.2</td>
</tr>
<tr>
<td></td>
<td>• AWS ENA driver 1.2.3.0</td>
</tr>
<tr>
<td></td>
<td>• Amazon EC2 Hibernate Agent 1.0 (rollback from 2.1.0 in the 2018.03.16 AMI release)</td>
</tr>
<tr>
<td></td>
<td>• AWS EC2WinUtilDriver 1.0.1 (for troubleshooting)</td>
</tr>
<tr>
<td></td>
<td><strong>Windows Server 2016</strong></td>
</tr>
<tr>
<td></td>
<td>• EC2Launch 1.3.2000080</td>
</tr>
<tr>
<td>2018.03.16</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>AWS has removed all Windows AMIs dated 2018.03.16 due to an issue with an unquoted path in the configuration for the Amazon EC2 Hibernate Agent. For more information, see Issue with the Hibernate Agent (2018.03.16 AMIs) (p. 110).</td>
</tr>
<tr>
<td>2018.03.06</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV driver 8.2.1</td>
</tr>
<tr>
<td>2018.02.23</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV driver 7.4.6 (rollback from 8.2 in the 2018.02.13 AMI release)</td>
</tr>
</tbody>
</table>
## Amazon Elastic Compute Cloud
### User Guide for Windows Instances
#### Details About AWS Windows AMI Versions

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018.02.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to February 13, 2018</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2400</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.2.160.0</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.225.1</td>
</tr>
<tr>
<td></td>
<td>• AWS PV driver 8.2</td>
</tr>
<tr>
<td></td>
<td>• AWS ENA driver 1.2.3.0</td>
</tr>
<tr>
<td></td>
<td>• AWS NVMe driver 1.0.0.146</td>
</tr>
<tr>
<td></td>
<td>• Amazon EC2 HibernateAgent 1.0.0</td>
</tr>
</tbody>
</table>

#### Windows Server 2016
- EC2Launch 1.3.740

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018.01.12</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to January 9, 2018</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018.01.05</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to January 2018</td>
</tr>
<tr>
<td></td>
<td>• Registry settings to enable mitigations for the Spectre and Meltdown exploits</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.215</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2262</td>
</tr>
</tbody>
</table>

### Monthly AMI Updates for 2017

For more information about Microsoft updates, see [Description of Software Update Services and Windows Server Update Services changes in content for 2017](#).

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017.12.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to December 12, 2017</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2218</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation templates 1.4.27</td>
</tr>
<tr>
<td></td>
<td>• AWS NVMe driver 1.02</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.2.93.0</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.201</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017.11.29</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Removed components for Volume Shadow Copy Service (VSS) included in 2017.11.18 and 2017.11.19 due to a compatibility issue with Windows Backup.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017.11.19</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• EC2 Hibernate Agent 1.0 (supports hibernation for Spot Instances)</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>2017.11.18</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to November 14, 2017</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2218</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.2.64.0</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.182</td>
</tr>
<tr>
<td></td>
<td>• Elastic Network Adapter (ENA) driver 1.0.8 (rollback from 1.2.2 in the</td>
</tr>
<tr>
<td></td>
<td>2017.10.13 AMI release)</td>
</tr>
<tr>
<td></td>
<td>• Query for the latest Windows AMI using Systems Manager Parameter Store</td>
</tr>
<tr>
<td><strong>Windows Server 2016</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• EC2Launch 1.3.640</td>
</tr>
<tr>
<td><strong>2017.10.13</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to October 11, 2017</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2188</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.2.30.0</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation templates 1.4.24</td>
</tr>
<tr>
<td></td>
<td>• Elastic Network Adapter (ENA) driver 1.2.2. (Windows Server 2008 R2</td>
</tr>
<tr>
<td></td>
<td>through Windows Server 2016)</td>
</tr>
<tr>
<td><strong>2017.10.04</strong></td>
<td><strong>Microsoft SQL Server</strong></td>
</tr>
<tr>
<td></td>
<td>Windows Server 2016 with Microsoft SQL Server 2017 AMIs are now public</td>
</tr>
<tr>
<td></td>
<td>in all regions.</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-SQL_2017_Enterprise-2017.10.04</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-SQL_2017_Standard-2017.10.04</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-SQL_2017_Web-2017.10.04</td>
</tr>
<tr>
<td></td>
<td>• Windows_Server-2016-English-Full-SQL_2017_Express-2017.10.04</td>
</tr>
<tr>
<td></td>
<td>Microsoft SQL Server 2017 supports the following features:</td>
</tr>
<tr>
<td></td>
<td>• Machine Learning Services with Python (ML and AI) and R language</td>
</tr>
<tr>
<td></td>
<td>support</td>
</tr>
<tr>
<td></td>
<td>• Automatic database tuning</td>
</tr>
<tr>
<td></td>
<td>• Clusterless Availability Groups</td>
</tr>
<tr>
<td></td>
<td>• Runs on Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server</td>
</tr>
<tr>
<td></td>
<td>(SLES), and Ubuntu. For more information, see the following Microsoft</td>
</tr>
<tr>
<td></td>
<td>article: Installation guidance for SQL Server on Linux. Not supported</td>
</tr>
<tr>
<td></td>
<td>on Amazon Linux.</td>
</tr>
<tr>
<td></td>
<td>• Windows-Linux cross-OS migrations</td>
</tr>
<tr>
<td></td>
<td>• Resumable online index rebuild</td>
</tr>
<tr>
<td></td>
<td>• Improved adaptive query processing</td>
</tr>
<tr>
<td></td>
<td>• Graph data support</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>2017.09.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to September 13, 2017</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2106</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.952.0</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.143</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation templates 1.4.21</td>
</tr>
<tr>
<td>2017.08.09</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to August 9, 2017</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.2016</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.879.0</td>
</tr>
<tr>
<td><strong>Windows Server 2012 R2</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Due to an internal error, these AMIs were released with an older version of AWS Tools for Windows PowerShell, 3.3.58.0.</td>
</tr>
<tr>
<td>2017.07.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to July 13, 2017</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.1981</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.847.0</td>
</tr>
<tr>
<td><strong>Windows Server 2016</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Intel SRIOV Driver 2.0.210.0</td>
</tr>
<tr>
<td>2017.06.14</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to June 14, 2017</td>
</tr>
<tr>
<td></td>
<td>• Updates for .NET Framework 4.7 installed from Windows Update</td>
</tr>
<tr>
<td></td>
<td>• Microsoft updates to address the &quot;privilege not held&quot; error using the PowerShell Stop-Computer cmdlet. For more information, see <a href="https://support.microsoft.com/en-us/help/4026641">Privilege not held error</a> on the Microsoft site.</td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.1900</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.805.0</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.99.0</td>
</tr>
<tr>
<td></td>
<td>• Internet Explorer 11 for the desktop is the default, instead of the immersive Internet Explorer</td>
</tr>
<tr>
<td><strong>Windows Server 2016</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• EC2Launch 1.3.610</td>
</tr>
<tr>
<td>2017.05.30</td>
<td>The Windows_Server-2008-SP2-English-32Bit-Base-2017.05.10 AMI was updated to the Windows_Server-2008-SP2-English-32Bit-Base-2017.05.30 AMI to resolve an issue with password generation.</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2017.05.22</td>
<td>The Windows_Server-2016-English-Full-Base-2017.05.10 AMI was updated to the Windows_Server-2016-English-Full-Base-2017.05.22 AMI after some log cleaning.</td>
</tr>
<tr>
<td>2017.05.10</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to May 9, 2017</td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver v7.4.6</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.83.0</td>
</tr>
<tr>
<td></td>
<td><strong>Windows Server 2016</strong></td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.767</td>
</tr>
<tr>
<td>2017.04.12</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to April 11, 2017</td>
</tr>
<tr>
<td></td>
<td>• AWS Tools for Windows PowerShell 3.3.71.0</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation templates 1.4.18</td>
</tr>
<tr>
<td></td>
<td><strong>Windows Server 2003 to Windows Server 2012</strong></td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.9.1775</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.761.0</td>
</tr>
<tr>
<td></td>
<td><strong>Windows Server 2016</strong></td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.730.0</td>
</tr>
<tr>
<td>2017.03.15</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to March 14, 2017</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Current AWS CloudFormation templates</td>
</tr>
<tr>
<td></td>
<td><strong>Windows Server 2003 to Windows Server 2012</strong></td>
</tr>
<tr>
<td></td>
<td>• EC2Config version 4.7.1631</td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.682.0</td>
</tr>
<tr>
<td></td>
<td><strong>Windows Server 2016</strong></td>
</tr>
<tr>
<td></td>
<td>• SSM Agent 2.0.706.0</td>
</tr>
<tr>
<td></td>
<td>• EC2Launch v1.3.540</td>
</tr>
<tr>
<td>2017.02.21</td>
<td>Microsoft recently announced that they will not release monthly patches or security updates for the month of February. All February patches and security updates will be included in the March update.</td>
</tr>
<tr>
<td></td>
<td>Amazon Web Services did not release updated Windows Server AMIs in February.</td>
</tr>
</tbody>
</table>
### Monthly AMI Updates for 2016

For more information about Microsoft updates, see Description of Software Update Services and Windows Server Update Services changes in content for 2016.

#### 2017.01.11

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>All AMIs</td>
<td>• Microsoft security updates current to January 10, 2017</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Current AWS CloudFormation templates</td>
</tr>
</tbody>
</table>

**Windows Server 2003 to Windows Server 2012**

- EC2Config version 4.2.1442
- SSM Agent 2.0.599.0

#### 2016.12.14

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>All AMIs</td>
<td>• Microsoft security updates current to December 13, 2016</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
</tbody>
</table>

**Windows Server 2003 to Windows Server 2012**

- Released EC2Config version 4.1.1396
- Elastic Network Adapter (ENA) driver 1.0.9.0 (Windows Server 2008 R2 only)

**Windows Server 2016**

- New AMIs available in all regions:
  - Windows_Server-2016-English-Core-Base

**Microsoft SQL Server**

All Microsoft SQL Server AMIs with the latest service pack are now public in all regions. These new AMIs replace old SQL Service Pack AMIs going forward.

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SQL Server 2016 SP1 is a major release. The following features, which were previously available in Enterprise edition only, are now enabled in Standard, Web, and Express editions with SQL Server 2016 SP1:</td>
</tr>
<tr>
<td></td>
<td>\begin{itemize}</td>
</tr>
<tr>
<td></td>
<td>\item Row-level security</td>
</tr>
<tr>
<td></td>
<td>\item Dynamic Data Masking</td>
</tr>
<tr>
<td></td>
<td>\item Change Data Capture</td>
</tr>
<tr>
<td></td>
<td>\item Database snapshot</td>
</tr>
<tr>
<td></td>
<td>\item Column store</td>
</tr>
<tr>
<td></td>
<td>\item Partitioning</td>
</tr>
<tr>
<td></td>
<td>\item Compression</td>
</tr>
<tr>
<td></td>
<td>\item In Memory OLTP</td>
</tr>
<tr>
<td></td>
<td>\item Always Encrypted</td>
</tr>
<tr>
<td></td>
<td>\end{itemize}</td>
</tr>
<tr>
<td>2016.11.23</td>
<td><strong>Windows Server 2003 to Windows Server 2012</strong></td>
</tr>
<tr>
<td></td>
<td>- Released EC2Config version 4.1.1378</td>
</tr>
<tr>
<td></td>
<td>- The AMIs released this month, and going forward, use the EC2Config service to process boot-time configurations and SSM Agent to process AWS Systems Manager Run Command and Config requests. EC2Config no longer processes requests for Systems Manager Run Command and State Manager. The latest EC2Config installer installs SSM Agent side-by-side with the EC2Config service. For more information, see \textit{EC2Config and AWS Systems Manager (p. 408)}.</td>
</tr>
<tr>
<td>2016.11.09</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to November 8 2016</td>
</tr>
<tr>
<td></td>
<td>- Released AWS PV driver, version 7.4.3.0 for Windows 2008 R2 and later</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2016.10.18</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to October 12, 2016</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td><strong>Windows Server 2016</strong></td>
</tr>
<tr>
<td></td>
<td>- Released AMIs for Windows Server 2016. These AMIs include significant changes. For example, they don't include the EC2Config service. For more information, see \textit{Changes in Windows Server 2016 and Later AMIs (p. 109)}.</td>
</tr>
<tr>
<td>2016.9.14</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to September 13, 2016</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>- Renamed AMI Windows_Server-2012-RTM-Japanese-64Bit-SQL_2008\textunderscore R3\textunderscore SP2\textunderscore Standard to Windows_Server-2012-RTM-Japanese-64Bit-SQL_2008\textunderscore R2\textunderscore SP3\textunderscore Standard</td>
</tr>
</tbody>
</table>

---

99
<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016.8.26</td>
<td>All Windows Server 2008 R2 AMIs dated 2016.08.11 were updated to fix a known issue. New AMIs are dated 2016.08.25.</td>
</tr>
</tbody>
</table>
| 2016.8.11 | **All AMIs**  
- Ec2Config v3.19.1153  
- Microsoft security updates current to August 10, 2016  
- Enabled the registry key User32 exception handler hardening feature in Internet Explorer for MS15-124 |
|           | - Elastic Network Adapter (ENA) Driver 1.0.8.0  
- ENA AMI property set to enabled  
- AWS PV Driver for Windows Server 2008 R2 was re-released this month because of a known issue. Windows Server 2008 R2 AMI's were removed in July because of this issue. |
| 2016.8.2  | All Windows Server 2008 R2 AMIs for July were removed and rolled back to AMIs dated 2016.06.15, because of an issue discovered in the AWS PV driver. The AWS PV driver issue has been fixed. The August AMI release will include Windows Server 2008 R2 AMIs with the fixed AWS PV driver and July/August Windows updates. |
| 2016.7.26 | **All AMIs**  
- Ec2Config v3.18.1118  
- 2016.07.13 AMIs were missing security patches. AMIs were re-patched. Additional processes were put in place to verify successful patch installations going forward. |
| 2016.7.13 | **All AMIs**  
- Microsoft security updates current to July 2016  
- Current AWS Tools for Windows PowerShell  
- Updated AWS PV Driver 7.4.2.0  
- AWS PV Driver for Windows Server 2008 R2 |
| 2016.6.16 | **All AMIs**  
- Microsoft security updates current to June 2016  
- Current AWS Tools for Windows PowerShell  
- EC2Config service version 3.17.1032 |
|           | **Microsoft SQL Server**                                                                                                           |
|           | - Released 10 AMIs that include 64-bit versions of Microsoft SQL Server 2016. If using the Amazon EC2 console, navigate to Images, AMIs, Public Images, and type Windows_Server-2012-R2_RTM-English-64Bit-Sql_2016_Standard in the search bar. For more information, see What's New in SQL Server 2016 on MSDN. |
## Details About AWS Windows AMI Versions

### Release Changes

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2016.5.11</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to May 2016</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>- EC2Config service version 3.16.930</td>
</tr>
<tr>
<td></td>
<td>- MS15-011 Active Directory patch installed</td>
</tr>
<tr>
<td><strong>Windows Server 2012 R2</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Intel SRIOV Driver 1.0.16.1</td>
</tr>
<tr>
<td><strong>2016.4.13</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to April 2016</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>- EC2Config service version 3.15.880</td>
</tr>
<tr>
<td><strong>2016.3.9</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to March 2016</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>- EC2Config service version 3.14.786</td>
</tr>
<tr>
<td><strong>2016.2.10</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to February 2016</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>- EC2Config service version 3.13.727</td>
</tr>
<tr>
<td><strong>2016.1.25</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to January 2016</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>- EC2Config service version 3.12.649</td>
</tr>
<tr>
<td><strong>2016.1.5</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
</tbody>
</table>

### Monthly AMI Updates for 2015

For more information about Microsoft updates, see Description of Software Update Services and Windows Server Update Services changes in content for 2015.

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2015.12.15</strong></td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>- Microsoft security updates current to December 2015</td>
</tr>
<tr>
<td></td>
<td>- Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2015.11.11</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to November 2015</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.11.521</td>
</tr>
<tr>
<td></td>
<td>• CFN Agent updated to latest version</td>
</tr>
<tr>
<td>2015.10.26</td>
<td>Corrected boot volume sizes of base AMIs to be 30GB instead of 35GB</td>
</tr>
<tr>
<td>2015.10.14</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to October 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.10.442</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Updated SQL Service Packs to latest versions for all SQL variants</td>
</tr>
<tr>
<td></td>
<td>• Removed old entries in Event Logs</td>
</tr>
<tr>
<td></td>
<td>• AMI Names have been changed to reflect the latest service pack. For example, the latest AMI with Server 2012 and SQL 2014 Standard is named &quot;Windows_Server-2012-RTM-English-64Bit-SQL_2014_SP1_Standard-2015.10.26&quot;, not &quot;Windows_Server-2012-RTM-English-64Bit-SQL_2014_RTM_Standard-2015.10.26&quot;.</td>
</tr>
<tr>
<td>2015.9.9</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to September 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.9.359</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Current AWS CloudFormation helper scripts</td>
</tr>
<tr>
<td>2015.8.18</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to August 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.8.294</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Windows Server 2012 and Windows Server 2012 R2</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver 7.3.2</td>
</tr>
<tr>
<td>2015.7.21</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to July 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.7.308</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Modified AMI descriptions of SQL images for consistency</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>2015.6.10</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to June 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.6.269</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Current AWS CloudFormation helper scripts</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Windows Server 2012 R2</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver 7.3.1</td>
</tr>
<tr>
<td>2015.5.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to May 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.5.228</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2015.04.15</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to April 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.3.174</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2015.03.11</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to March 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.2.97</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Windows Server 2012 R2</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver 7.3.0</td>
</tr>
<tr>
<td>2015.02.11</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to February 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 3.0.54</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Current AWS CloudFormation helper scripts</td>
</tr>
<tr>
<td>2015.01.14</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to January 2015</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.3.313</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Current AWS CloudFormation helper scripts</td>
</tr>
</tbody>
</table>
## Monthly AMI Updates for 2014

For more information about Microsoft updates, see Description of Software Update Services and Windows Server Update Services changes in content for 2014.

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014.12.10</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to December 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.12</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2014.11.19</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to November 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.11</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2014.10.15</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to October 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.10</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Windows Server 2012 R2</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver 7.2.4.1 (resolves the issues with Plug and Play Cleanup, which is now enabled by default)</td>
</tr>
<tr>
<td>2014.09.10</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to September 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.8</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Windows Server 2012 R2</strong></td>
</tr>
<tr>
<td></td>
<td>• Disable Plug and Play Cleanup (see Important information)</td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver 7.2.2.1 (resolves issues with the uninstaller)</td>
</tr>
<tr>
<td>2014.08.13</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to August 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.7</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Windows Server 2012 R2</strong></td>
</tr>
<tr>
<td></td>
<td>• AWS PV Driver 7.2.2.1 (improves disk performance, resolves issues with reconnecting multiple network interfaces and lost network settings)</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2014.07.10</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to July 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.5</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2014.06.12</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to June 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.4</td>
</tr>
<tr>
<td></td>
<td>• Removed NVIDIA drivers (except for Windows Server 2012 R2 AMIs)</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2014.05.14</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to May 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.2</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation helper scripts version 1.4.0</td>
</tr>
<tr>
<td>2014.04.09</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to April 2014</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Current AWS CloudFormation helper scripts</td>
</tr>
<tr>
<td>2014.03.12</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to March 2014</td>
</tr>
<tr>
<td>2014.02.12</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to February 2014</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.2.1</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• KB2634328</td>
</tr>
<tr>
<td></td>
<td>• Remove the BCDEdit useplatformclock value</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Microsoft SQL Server</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft SQL Server 2012 SP1 cumulative update package 8</td>
</tr>
<tr>
<td></td>
<td>• Microsoft SQL Server 2008 R2 cumulative update package 10</td>
</tr>
</tbody>
</table>

**Monthly AMI Updates for 2013**

For more information about Microsoft updates, see [Description of Software Update Services and Windows Server Update Services changes in content for 2013](#).
<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013.11.13</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to November 2013</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.1.19</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• Configure NTP to synchronize the time once a day (the default is every seven days)</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Windows Server 2012</strong></td>
</tr>
<tr>
<td></td>
<td>• Clean up the WinSXS folder using the following command: _dism online _ cleanup-image <em>/StartComponentCleanup</em></td>
</tr>
<tr>
<td>2013.09.11</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to September 2013</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.1.18</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation helper scripts version 1.3.15</td>
</tr>
<tr>
<td>2013.07.10</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to July 2013</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.1.16</td>
</tr>
<tr>
<td></td>
<td>• Expanded the root volume to 50 GB</td>
</tr>
<tr>
<td></td>
<td>• Set the page file to 512 MB, expanding to 8 GB as needed</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2013.06.12</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to June 2013</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td><strong>Only AMIs with Microsoft SQL Server</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft SQL Server 2012 SP1 with cumulative update package 4</td>
</tr>
<tr>
<td>2013.05.15</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to May 2013</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.1.15</td>
</tr>
<tr>
<td></td>
<td>• All instance store volumes attached by default</td>
</tr>
<tr>
<td></td>
<td>• Remote PowerShell enabled by default</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td>2013.04.14</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to April 2013</td>
</tr>
<tr>
<td></td>
<td>• Current AWS Tools for Windows PowerShell</td>
</tr>
<tr>
<td></td>
<td>• AWS CloudFormation helper scripts version 1.3.14</td>
</tr>
</tbody>
</table>
## Monthly AMI Updates for 2012

For more information about Microsoft updates, see Description of Software Update Services and Windows Server Update Services changes in content for 2012.

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012.12.12</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to December 2012</td>
</tr>
<tr>
<td></td>
<td>• Set the ActiveTimeBias registry value to 0</td>
</tr>
<tr>
<td></td>
<td>• Disable IPv6 for the network adapter</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.1.9</td>
</tr>
<tr>
<td></td>
<td>• Add AWS Tools for Windows PowerShell and set the policy to allow import-module</td>
</tr>
<tr>
<td>2012.11.15</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to November 2012</td>
</tr>
<tr>
<td></td>
<td>• EC2Config service version 2.1.7</td>
</tr>
<tr>
<td>2012.10.10</td>
<td><strong>All AMIs</strong></td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to October 2012</td>
</tr>
<tr>
<td>Release</td>
<td>Changes</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2012.08.15</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to August 2012&lt;br&gt;• EC2Config service version 2.1.2&lt;br&gt;• KB2545227</td>
</tr>
<tr>
<td>2012.07.11</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to July 2012</td>
</tr>
<tr>
<td>2012.06.12</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to June 2012&lt;br&gt;• Set page file to 4 GB&lt;br&gt;• Remove installed language packs&lt;br&gt;• Set performance option to &quot;Adjust for best performance&quot;&lt;br&gt;• Set the screen saver to no longer display the logon screen on resume&lt;br&gt;• Remove previous RedHat driver versions using pnputil&lt;br&gt;• Remove duplicate bootloaders and set bootstatuspolicy to ignoreallfailures using bcdedit</td>
</tr>
<tr>
<td>2012.05.10</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to May 2012&lt;br&gt;• EC2Config service version 2.1.0</td>
</tr>
<tr>
<td>2012.04.11</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to April 2012&lt;br&gt;• KB2582281&lt;br&gt;• Current version of EC2Config&lt;br&gt;• System time in UTC instead of GMT</td>
</tr>
<tr>
<td>2012.03.13</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to March 2012</td>
</tr>
<tr>
<td>2012.02.24</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to February 2012&lt;br&gt;• Standardize AMI names and descriptions</td>
</tr>
<tr>
<td>2012.01.12</td>
<td><strong>All AMIs</strong>&lt;br&gt;• Microsoft security updates current to January 2012&lt;br&gt;• RedHat PV driver version 1.3.10</td>
</tr>
</tbody>
</table>
Monthly AMI Updates for 2011 and earlier

<table>
<thead>
<tr>
<th>Release</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011.09.11</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Microsoft security updates current to September 2011</td>
</tr>
<tr>
<td>1.04</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Current Microsoft security updates</td>
</tr>
<tr>
<td></td>
<td>• Update network driver</td>
</tr>
<tr>
<td></td>
<td>• Fix issue with instances in a VPC losing connectivity when changing the time zone of the instance</td>
</tr>
<tr>
<td>1.02</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Current Microsoft security updates</td>
</tr>
<tr>
<td></td>
<td>• Update network driver</td>
</tr>
<tr>
<td></td>
<td>• Add support for licensing activation for instances in a VPC</td>
</tr>
<tr>
<td>1.01</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Current Microsoft security updates</td>
</tr>
<tr>
<td></td>
<td>• Fix issue with password improperly generated while waiting for network availability</td>
</tr>
<tr>
<td>1.0</td>
<td>All AMIs</td>
</tr>
<tr>
<td></td>
<td>• Initial release</td>
</tr>
</tbody>
</table>

Changes in Windows Server 2016 and Later AMIs

AWS provides AMIs for Windows Server 2016 and later. These AMIs include the following high-level changes from earlier Windows AMIs:

- To accommodate the change from .NET Framework to .NET Core, the EC2Config service has been deprecated on Windows Server 2016 AMIs and replaced by EC2Launch. EC2Launch is a bundle of Windows PowerShell scripts that perform many of the tasks performed by the EC2Config service. For more information, see Configuring a Windows Instance Using EC2Launch (p. 395).

- On earlier versions of Windows Server AMIs, you can use the EC2Config service to join an EC2 instance to a domain and configure integration with Amazon CloudWatch. On Windows Server 2016 and later AMIs, you can use the CloudWatch agent to configure integration with Amazon CloudWatch. For more information about configuring instances to send log data to CloudWatch, see Collect Metrics and Logs from Amazon EC2 Instances and On-Premises Servers with the CloudWatch Agent. For information about joining an EC2 instance to a domain, see Join an Instance to a Domain Using the AWS-JoinDirectoryServiceDomain JSON Document in the AWS Systems Manager User Guide.

Other Differences

Note these additional important differences for instances created from Windows Server 2016 and later AMIs.
• By default, EC2Launch does not initialize secondary EBS volumes. You can configure EC2Launch to initialize disks automatically by either scheduling the script to run or by calling EC2Launch in user data. For the procedure to initialize disks using EC2Launch, see "Initialize Drives and Drive Letter Mappings" in Configuring EC2Launch (p. 397).
• If you previously enabled CloudWatch integration on your instances by using a local configuration file (AWS.EC2.Windows.CloudWatch.json), you can configure the file to work with the SSM Agent on instances created from Windows Server 2016 and later AMIs.

For more information, see Windows Server 2019 on Microsoft.com.

Docker Container Conflict on Windows Server 2016 Instances

If you run the Docker service on Windows Server 2016 AMIs, the service is configured to use a different CIDR value than the default internal IP address prefix value. The default value is 172.16.0.0/12. Windows Server 2016 AMIs use 172.17.0.0/16 to avoid conflict with the default Amazon EC2 VPC/subnet. If you don’t change VPC/subnet settings for your EC2 instances, then you don’t need to do anything. The conflict is essentially avoided because of the different CIDR values. If you do change VPC/subnet settings, be aware of these internal IP address prefix values and avoid creating a conflict. For more information, read the following section.

Important
If you plan to run Docker on a Windows Server 2016 instance, you must create the instance from the following Amazon Machine Image (AMI) or an AMI based on an image with Windows_Server-2016-English-Full-Containers in the name. Otherwise, if you use a different Windows Server 2016 AMI, instances fail to boot correctly after installing Docker and then running Sysprep.

Issue with the Hibernate Agent (2018.03.16 AMIs)

After the release of the 2018.03.16 Windows AMIs, we discovered an unquoted path in the configuration of the Amazon EC2 Hibernate Agent. The agent was included in the AMIs for Windows Server 2008 through Windows Server 2016. This issue does not impact the AMIs for Windows Server 2003.

AWS has removed the Windows AMIs dated 2018.03.16. To be notified when new Windows AMIs are available, see Subscribing to Windows AMI Notifications (p. 79).

To mitigate the issue, you can use one of the following procedures to add the missing quotation marks. If the agent is running, you must also restart the agent. Alternatively, you can terminate any instances that you launched from a 2018.03.16 Windows AMI and replace them with instances launched using a different AMI.

Windows PowerShell

2. Use the following command to update the configuration, adding the missing quotation marks:

```cmd /c 'sc config EC2HibernateAgent binPath=""%ProgramFiles%\Amazon\Hibernate\EC2HibernateAgent.exe""'
```

3. Use the following command to view the updated configuration:

```(Get-ItemProperty -Path Registry::HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\EC2HibernateAgent\).ImagePath```
Verify that the response is enclosed in quotation marks, as shown in the following example:

"C:\Program Files\Amazon\Hibernate\EC2HibernateAgent.exe"

4. Use the following command to check whether Status is Running:

```powershell
Get-Service EC2HibernateAgent
```

If the agent is running, you must restart it using the following command so that the change takes effect:

```powershell
Restart-Service EC2HibernateAgent
```

Command Prompt

1. On your Windows instance, open a Command Prompt window.
2. Use the following command to update the configuration, adding the missing quotation marks:

```cmd
sc config EC2HibernateAgent binPath="%ProgramFiles%\Amazon\Hibernate\EC2HibernateAgent.exe"
```

3. Use the following command to view the updated configuration:

```cmd
sc qc EC2HibernateAgent
```

Verify that the path in BINARY_PATH_NAME is enclosed in quotation marks, as shown in the following example:

"C:\Program Files\Amazon\Hibernate\EC2HibernateAgent.exe"

4. Use the following command to check whether STATE is RUNNING:

```cmd
sc query EC2HibernateAgent
```

If the agent is running, you must restart it using the following command so that the change takes effect:

```cmd
sc stop EC2HibernateAgent && sc start EC2HibernateAgent
```

Amazon EC2 Windows Server AMIs for STIG Compliance

Security Technical Implementation Guides (STIGs) are the configuration standards created by the Defense Information Systems Agency (DISA) to secure information systems and software. To make your systems compliant with STIG standards, you must install, configure, and test a variety of security settings. Amazon EC2 Windows Server AMIs for STIG Compliance are pre-configured with over 160 required security settings. STIG-compliant operating systems include Windows Server 2012 R2, Windows Server 2016, and Windows Server 2019. The STIG-compliant AMIs include updated Department of Defense (DoD) certificates to help you get started and achieve STIG compliance. There are no additional charges for using STIG-compliant AMIs.
Amazon EC2 Windows Server AMIs for STIG compliance are available in all public AWS and GovCloud regions. You can launch instances from these AMIs directly from the Amazon EC2 console. They are billed using standard Windows pricing.

The following lists and tables show the STIGs that have not been applied to each Operating System, by category, unless otherwise indicated.

**Topics**
- Core and Base Operating System (p. 112)
- Internet Explorer (IE) 11 Using STIG V1 Release 17 (p. 112)
- Microsoft .NET Framework 4.0 Using STIG V1 Release 8: STIGS Not Applied (p. 112)

**Core and Base Operating System**

The following STIG settings have not been applied. All other applicable STIGs have been applied. For complete list, see https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_MS_Windows_2012_and_2012_R2_MS_V2R16_STIG.zip.

**Windows Server 2012 R2 STIG V2 Release 16**


**Windows Server 2016 STIG V1 Release 9**


**Windows Server 2019 STIG V1 Release 2**


**Internet Explorer (IE) 11 Using STIG V1 Release 17**

All STIG settings related to Internet Explorer have been applied. For complete list, see https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_MS_IE11_V1R17_STIG.zip.

**Windows Server 2012 R2, 2016, and 2019**

All STIG settings applied.

**Microsoft .NET Framework 4.0 Using STIG V1 Release 8: STIGS Not Applied**

The following STIGs have not been applied. All other applicable STIG settings have been applied. For complete list, see https://dl.dod.cyber.mil/wp-content/uploads/stigs/zip/U_MS_DotNet_Framework_4-0_V1R8_STIG.zip.

**Windows Server 2012 R2, 2016, and 2019**

112
Create a Standard Amazon Machine Image Using Sysprep

The Microsoft System Preparation (Sysprep) tool simplifies the process of duplicating a customized installation of Windows. We recommend that you use Sysprep to create a standardized Amazon Machine Image (AMI). You can then create new Amazon EC2 instances for Windows from this standardized image.

We also recommend that you run Sysprep with EC2Launch (Windows Server 2016 and later) or the EC2Config service (prior to Windows Server 2016).

Important

Don't use Sysprep to create an instance backup. Sysprep removes system-specific information; removing this information might have unintended consequences for an instance backup.

Contents
- Before You Begin (p. 113)
- Using Sysprep with the EC2Config Service (p. 113)
- Run Sysprep with the EC2Config Service (p. 116)
- Troubleshooting Sysprep (p. 117)

Before You Begin

- Before performing Sysprep, we recommend that you remove all local user accounts and all account profiles other than a single administrator account under which Sysprep will be executed. If you perform Sysprep with additional accounts and profiles, unexpected behavior could result, including loss of profile data or failure to complete Sysprep.
- Learn more about Sysprep on Microsoft TechNet.
- Learn which server roles are supported for Sysprep.
- The procedures on this page apply to E2Config. With Windows Server 2016 and later, see Using Sysprep with EC2Launch (p. 400).

Using Sysprep with the EC2Config Service

Learn the details of the different Sysprep execution phases and the tasks performed by the EC2Config service as the image is prepared.

Sysprep Phases

Sysprep runs through the following phases:

- **Generalize**: The tool removes image-specific information and configurations. For example, Sysprep removes the security identifier (SID), the computer name, the event logs, and specific drivers, to name a few. After this phase is completed, the operating system (OS) is ready to create an AMI.

  **Note**

  When you run Sysprep with the EC2Config service, the system prevents drivers from being removed because the PersistAllDeviceInstalls setting is set to true by default.
• **Specialize**: Plug and Play scans the computer and installs drivers for any detected devices. The tool generates OS requirements like the computer name and SID. Optionally, you can execute commands in this phase.

• **Out-of-Box Experience (OOBE)**: The system runs an abbreviated version of Windows Setup and asks the user to enter information such as a system language, the time zone, and a registered organization. When you run Sysprep with EC2Config, the answer file automates this phase.

### Sysprep Actions

Sysprep and the EC2Config service perform the following actions when preparing an image.

1. When you choose **Shutdown with Sysprep** in the **EC2 Service Properties** dialog box, the system runs the `ec2config.exe --sysprep` command.

2. The EC2Config service reads the content of the `BundleConfig.xml` file. This file is located in the following directory, by default: `C:\Program Files\Amazon\Ec2ConfigService\Settings`.

   The `BundleConfig.xml` file includes the following settings. You can change these settings:
   - **AutoSysprep**: Indicates whether to use Sysprep automatically. You do not need to change this value if you are running Sysprep from the EC2 Service Properties dialog box. The default value is No.
   - **SetRDPCertificate**: Sets a self-signed certificate for the Remote Desktop server. This enables you to securely use the Remote Desktop Protocol (RDP) to connect to the instance. Change the value to Yes if new instances should use a certificate. This setting is not used with Windows Server 2008 or Windows Server 2012 instances because these operating systems can generate their own certificates. The default value is No.
   - **SetPasswordAfterSysprep**: Sets a random password on a newly launched instance, encrypts it with the user launch key, and outputs the encrypted password to the console. Change the value to No if new instances should not be set to a random encrypted password. The default value is Yes.
   - **PreSysprepRunCmd**: The location of the command to run. The command is located in the following directory, by default: `C:\Program Files\Amazon\Ec2ConfigService\Scripts\BeforeSysprep.cmd`.

3. The system executes `BeforeSysprep.cmd`. This command creates a registry key as follows:

   ```
   reg add "HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Terminal Server" /v fDenyTSConnections /t REG_DWORD /d 1 /f
   ```

   The registry key disables RDP connections until they are re-enabled. Disabling RDP connections is a necessary security measure because, during the first boot session after Sysprep has run, there is a short period of time where RDP allows connections and the Administrator password is blank.

4. The EC2Config service calls Sysprep by running the following command:

   ```
   sysprep.exe /unattend: "C:\Program Files\Amazon\Ec2ConfigService\sysprep2008.xml" /oobe /generalize /shutdown
   ```

### Generalize Phase

The tool removes image-specific information and configurations such as the computer name and the SID. If the instance is a member of a domain, it is removed from the domain. The `sysprep2008.xml` answer file includes the following settings which affect this phase:

- **PersistAllDeviceInstalls**: This setting prevents Windows Setup from removing and reconfiguring devices, which speeds up the image preparation process because Amazon AMIs require certain drivers to run and re-detection of those drivers would take time.

- **DoNotCleanUpNonPresentDevices**: This setting retains Plug and Play information for devices that are not currently present.
• Sysprep shuts down the OS as it prepares to create the AMI. The system either launches a new instance or starts the original instance.

Specialize Phase

The system generates OS specific requirements such as a computer name and a SID. The system also performs the following actions based on configurations that you specify in the sysprep2008.xml answer file.

• CopyProfile: Sysprep can be configured to delete all user profiles, including the built-in Administrator profile. This setting retains the built-in Administrator account so that any customizations you made to that account are carried over to the new image. The default value is True.

CopyProfile replaces the default profile with the existing local administrator profile. All accounts logged into after running Sysprep will receive a copy of that profile and its contents at first login.

If you don’t have specific user-profile customizations that you want to carry over to the new image then change this setting to False. Sysprep will remove all user profiles; this saves time and disk space.

• TimeZone: The time zone is set to Coordinate Universal Time (UTC) by default.

• Synchronous command with order 1: The system executes the following command that enables the administrator account and specifies the password requirement.

```
net user Administrator /ACTIVE:YES /LOGONPASSWORDCHG:NO /EXPIRES:NEVER /PASSWORDREQ:YES
```

• Synchronous command with order 2: The system scrambles the administrator password. This security measure is designed to prevent the instance from being accessible after Sysprep completes if you did not enable the ec2setpassword setting.

```
C:\Program Files\Amazon\Ec2ConfigService\ScramblePassword.exe -u Administrator
```

• Synchronous command with order 3: The system executes the following command:

```
C:\Program Files\Amazon\Ec2ConfigService\Scripts\SysprepSpecializePhase.cmd
```

This command adds the following registry key, which re-enables RDP:

```
reg add "HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\Terminal Server" /v fDenyTSCConnections /t REG_DWORD /d 0 /f
```

OOBE Phase

1. Using the EC2Config service answer file, the system specifies the following configurations:

   • <InputLocale>en-US</InputLocale>
   • <SystemLocale>en-US</SystemLocale>
   • <UILanguage>en-US</UILanguage>
   • <UserLocale>en-US</UserLocale>
   • <HideEULAPage>true</HideEULAPage>
   • <HideWirelessSetupInOOBE>true</HideWirelessSetupInOOBE>
   • <NetworkLocation>Other</NetworkLocation>
   • <ProtectYourPC>3</ProtectYourPC>
   • <BluetoothTaskbarIconEnabled>false</BluetoothTaskbarIconEnabled>
   • <TimeZone>UTC</TimeZone>
   • <RegisteredOrganization>Amazon.com</RegisteredOrganization>
   • <RegisteredOwner>Amazon</RegisteredOwner>
Note
During the generalize and specialize phases the EC2Config service monitors the status of the OS. If EC2Config detects that the OS is in a Sysprep phase, then it publishes the following message to the system log:

EC2ConfigMonitorState: 0 Windows is being configured. SysprepState=IMAGE_STATE_UNDEPLOYABLE

2. After the OOBE phase completes, the system executes the SetupComplete.cmd from the following location: C:\Windows\Setup\Scripts\SetupComplete.cmd. In Amazon public AMIs before April 2015 this file was empty and executed nothing on the image. In public AMIs dated after April 2015, the file includes the following value: call "C:\Program Files\Amazon\Ec2ConfigService\Scripts\PostSysprep.cmd".

3. The system executes the PostSysprep.cmd, which performs the following operations:
   - Sets the local Administrator password to not expire. If the password expired, Administrators might not be able to log on.
   - Sets the MSSQLServer machine name (if installed) so that the name will be in sync with the AMI.

Post Sysprep

After Sysprep completes, the EC2Config services sends the following message to the console output:

<table>
<thead>
<tr>
<th>Windows sysprep configuration complete.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Message: Sysprep Start</td>
</tr>
<tr>
<td>Message: Sysprep End</td>
</tr>
</tbody>
</table>

EC2Config then performs the following actions:

1. Reads the content of the config.xml file and lists all enabled plug-ins.
2. Executes all “Before Windows is ready” plug-ins at the same time.
   - Ec2SetPassword
   - Ec2SetComputerName
   - Ec2InitializeDrives
   - Ec2EventLog
   - Ec2ConfigureRDP
   - Ec2OutputRDP Cert
   - Ec2SetDriveLetter
   - Ec2WindowsActivate
   - Ec2DynamicBootVolumeSize
3. After it is finished, sends a “Windows is ready” message to the instance system logs.
4. Runs all “After Windows is ready” plug-ins at the same time.
   - AWS CloudWatch logs
   - UserData
   - AWS Systems Manager (Systems Manager)

For more information about Windows plug-ins, see Configuring a Windows Instance Using the EC2Config Service (p. 405).

Run Sysprep with the EC2Config Service

Use the following procedure to create a standardized AMI using Sysprep and the EC2Config service.
1. In the Amazon EC2 console locate or create (p. 64) an AMI that you want to duplicate.
2. Launch and connect to your Windows instance.
3. Customize it.
4. Specify configuration settings in the EC2Config service answer file:
   
   C:\Program Files\Amazon\Ec2ConfigService\sysprep2008.xml

5. From the Windows Start menu, choose All Programs, and then choose EC2ConfigService Settings.
6. Choose the Image tab in the Ec2 Service Properties dialog box. For more information about the options and settings in the Ec2 Service Properties dialog box, see Ec2 Service Properties (p. 405).
7. Select an option for the Administrator password, and then select Shutdown with Sysprep or Shutdown without Sysprep. EC2Config edits the settings files based on the password option that you selected.
   
   • Random: EC2Config generates a password, encrypts it with user's key, and displays the encrypted password to the console. We disable this setting after the first launch so that this password persists if the instance is rebooted or stopped and started.
   
   • Specify: The password is stored in the Sysprep answer file in unencrypted form (clear text). When Sysprep runs next, it sets the Administrator password. If you shut down now, the password is set immediately. When the service starts again, the Administrator password is removed. It's important to remember this password, as you can't retrieve it later.
   
   • Keep Existing: The existing password for the Administrator account doesn't change when Sysprep is run or EC2Config is restarted. It's important to remember this password, as you can't retrieve it later.

8. Choose OK.

When you are asked to confirm that you want to run Sysprep and shut down the instance, click Yes. You'll notice that EC2Config runs Sysprep. Next, you are logged off the instance, and the instance is shut down. If you check the Instances page in the Amazon EC2 console, the instance state changes from running to stopping, and then finally to stopped. At this point, it's safe to create an AMI from this instance.

You can manually invoke the Sysprep tool from the command line using the following command:

```
"%programfiles%\amazon\ec2configservice\"ec2config.exe -sysprep"
```

Note
The double quotation marks in the command are not required if your CMD shell is already in the C:\Program Files\Amazon\EC2ConfigService\ directory.

However, you must be very careful that the XML file options specified in the Ec2ConfigService Settings folder are correct; otherwise, you might not be able to connect to the instance. For more information about the settings files, see EC2Config Settings Files (p. 411). For an example of configuring and then running Sysprep from the command line, see Ec2ConfigService\Scripts\InstallUpdates.ps1.

Troubleshooting Sysprep

If you experience problems or receive error messages during image preparations, review the following logs:

- %WINDIR%\Panther\Unattendgc
- %WINDIR%\System32\Sysprep\Panther
- "C:\Program Files\Amazon\Ec2ConfigService\Logs\Ec2ConfigLog.txt"
If you receive an error message during image preparation with Sysprep, the OS might not be reachable. To review the log files, you must stop the instance, attach its root volume to another healthy instance as a secondary volume, and then review the logs mentioned earlier on the secondary volume.

If you locate errors in the Unattendgc log file, use the Microsoft Error Lookup Tool to get more details about the error. The following issue reported in the Unattendgc log file is typically the result of one or more corrupted user profiles on the instance:

```
Error [Shell Unattend] _FindLatestProfile failed (0x80070003) [gle=0x00000003]
Error [Shell Unattend] CopyProfile failed (0x80070003) [gle=0x00000003]
```

There are two options for resolving this issue:

**Option 1:** Use Regedit on the instance to search for the following key. Verify that there are no profile registry keys for a deleted user:

[HKEY_LOCAL_MACHINE\Software\Microsoft\Windows NT\CurrentVersion\ProfileList]

**Option 2:** Edit the EC2Config answer file (C:\Program Files\Amazon\Ec2ConfigService\sysprep2008.xml) and change `<CopyProfile>true</CopyProfile>` to `<CopyProfile>false</CopyProfile>`. Run Sysprep again. Note that this configuration change will delete the built-in administrator user profile after Sysprep completes.
Amazon EC2 Instances

If you're new to Amazon EC2, see the following topics to get started:

- What Is Amazon EC2? (p. 1)
- Setting Up with Amazon EC2 (p. 12)
- Getting Started with Amazon EC2 Windows Instances (p. 20)
- Instance Lifecycle (p. 329)

Before you launch a production environment, you need to answer the following questions.

Q. What instance type best meets my needs?

Amazon EC2 provides different instance types to enable you to choose the CPU, memory, storage, and networking capacity that you need to run your applications. For more information, see Instance Types (p. 119).

Q. What purchasing option best meets my needs?

Amazon EC2 supports On-Demand Instances (the default), Spot Instances, and Reserved Instances. For more information, see Instance Purchasing Options (p. 191).

Q. Can I remotely manage a fleet of EC2 instances and machines in my hybrid environment?

AWS Systems Manager enables you to remotely and securely manage the configuration of your Amazon EC2 instances, and your on-premises instances and virtual machines (VMs) in hybrid environments, including VMs from other cloud providers. For more information, see the AWS Systems Manager User Guide.

Instance Types

When you launch an instance, the instance type that you specify determines the hardware of the host computer used for your instance. Each instance type offers different compute, memory, and storage capabilities and are grouped in instance families based on these capabilities. Select an instance type based on the requirements of the application or software that you plan to run on your instance.

Amazon EC2 provides each instance with a consistent and predictable amount of CPU capacity, regardless of its underlying hardware.

Amazon EC2 dedicates some resources of the host computer, such as CPU, memory, and instance storage, to a particular instance. Amazon EC2 shares other resources of the host computer, such as the network and the disk subsystem, among instances. If each instance on a host computer tries to use as much of one of these shared resources as possible, each receives an equal share of that resource. However, when a resource is underused, an instance can consume a higher share of that resource while it's available.

Each instance type provides higher or lower minimum performance from a shared resource. For example, instance types with high I/O performance have a larger allocation of shared resources. Allocating a larger share of shared resources also reduces the variance of I/O performance. For most applications, moderate I/O performance is more than enough. However, for applications that require greater or more consistent I/O performance, consider an instance type with higher I/O performance.

Contents

- Available Instance Types (p. 120)
- Hardware Specifications (p. 121)
Available Instance Types

Amazon EC2 provides the instance types listed in the following tables.

## Current Generation Instances

For the best performance, we recommend that you use the current generation instance types when you launch new instances.

For more information about the current generation instance types, see [Amazon EC2 Instance Types](https://docs.aws.amazon.com/AmazonEC2/latest/UserGuide/InstanceTypes.html).

<table>
<thead>
<tr>
<th>Instance Family</th>
<th>Current Generation Instance Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>General purpose</td>
<td>m4.large</td>
</tr>
<tr>
<td>Compute optimized</td>
<td>r4.large</td>
</tr>
<tr>
<td>Memory optimized</td>
<td>r5.ad.large</td>
</tr>
</tbody>
</table>
## Hardware Specifications

For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.
To determine which instance type best meets your needs, we recommend that you launch an instance and use your own benchmark application. Because you pay by the instance hour, it’s convenient and inexpensive to test multiple instance types before making a decision.

If your needs change, even after you make a decision, you can resize your instance later. For more information, see Changing the Instance Type (p. 188).

**Note**

Amazon EC2 instances typically run on 64-bit virtual Intel processors as specified in the instance type product pages. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types. However, confusion may result from industry naming conventions for 64-bit CPUs. Chip manufacturer Advanced Micro Devices (AMD) introduced the first commercially successful 64-bit architecture based on the Intel x86 instruction set. Consequently, the architecture is widely referred to as AMD64 regardless of the chip manufacturer. Windows and several Linux distributions follow this practice. This explains why the internal system information on an Ubuntu or Windows EC2 instance displays the CPU architecture as AMD64 even though the instances are running on Intel hardware.

**Nitro-based Instances**

The Nitro system is a collection of AWS-built hardware and software components that enable high performance, high availability, and high security. In addition, the Nitro system provides bare metal capabilities that eliminate virtualization overhead and support workloads that require full access to host hardware.

**Nitro Components**

The following components are part of the Nitro system:

- Nitro hypervisor - A lightweight hypervisor that manages memory and CPU allocation and delivers performance that is indistinguishable from bare metal for most workloads.
- Nitro card
  - Local NVMe storage volumes
  - Networking hardware support
  - Management
  - Monitoring
  - Security
- Nitro security chip, integrated into the motherboard

**Instance Types**

The following instances are based on the Nitro system:

- C5, C5d, C5n, G4, I3en, M5, M5a, M5ad, M5d, p3dn.24xlarge, R5, R5a, R5ad, R5d, T3, T3a, and z1d
- Bare metal: c5.metal, c5n.metal, i3.metal, i3en.metal, m5.metal, m5d.metal, r5.metal, r5d.metal, u-6tb1.metal, u-9tb1.metal, u-12tb1.metal, and z1d.metal

**Resources**

For more information, see the following videos:

- AWS re:Invent 2017: The Amazon EC2 Nitro System Architecture
- AWS re:Invent 2017: Amazon EC2 Bare Metal Instances
- The Nitro Project: Next-Generation EC2 Infrastructure
Networking and Storage Features

When you select an instance type, this determines the networking and storage features that are available.

Networking features

- IPv6 is supported on all current generation instance types and the C3, R3, and I2 previous generation instance types.
- To maximize the networking and bandwidth performance of your instance type, you can do the following:
  - Launch supported instance types into a cluster placement group to optimize your instances for high performance computing (HPC) applications. Instances in a common cluster placement group can benefit from high-bandwidth, low-latency networking. For more information, see Placement Groups (p. 710).
  - Enable enhanced networking for supported current generation instance types to get significantly higher packet per second (PPS) performance, lower network jitter, and lower latencies. For more information, see Enhanced Networking on Windows (p. 700).
- Current generation instance types that are enabled for enhanced networking have the following networking performance attributes:
  - Traffic within the same region over private IPv4 or IPv6 can support 5 Gbps for single-flow traffic and up to 25 Gbps for multi-flow traffic (depending on the instance type).
  - Traffic to and from Amazon S3 buckets within the same region over the public IP address space or through a VPC endpoint can use all available instance aggregate bandwidth.
  - The maximum supported MTU varies across instance types. All Amazon EC2 instance types support standard Ethernet V2 1500 MTU frames. All current generation instances support 9001 MTU, or jumbo frames, and some previous generation instances support them as well. For more information, see Network Maximum Transmission Unit (MTU) for Your EC2 Instance (p. 718).

Storage features

- Some instance types support EBS volumes and instance store volumes, while other instance types support only EBS volumes. Some instance types that support instance store volumes use solid state drives (SSD) to deliver very high random I/O performance. Some instance types support NVMe instance store volumes. Some instance types support NVMe EBS volumes. For more information, see Amazon EBS and NVMe on Windows Instances (p. 875) and NVMe SSD Volumes (p. 925).
- To obtain additional, dedicated capacity for Amazon EBS I/O, you can launch some instance types as EBS–optimized instances. Some instance types are EBS–optimized by default. For more information, see Amazon EBS–Optimized Instances (p. 876).

Summary of Networking and Storage Features

The following table summarizes the networking and storage features supported by current generation instance types.

<table>
<thead>
<tr>
<th></th>
<th>EBS only</th>
<th>NVMe EBS</th>
<th>Instance store</th>
<th>Placement group</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Intel 82599 VF</td>
</tr>
<tr>
<td>C5</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>C5d</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
</tbody>
</table>
## Networking and Storage Features

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>EBS only</th>
<th>NVMe EBS</th>
<th>Instance store</th>
<th>Placement group</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5n</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>D2</td>
<td>No</td>
<td>No</td>
<td>HDD</td>
<td>Yes</td>
<td>Intel 82599 VF</td>
</tr>
<tr>
<td>F1</td>
<td>No</td>
<td>No</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>G3</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>G4</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>HS1</td>
<td>No</td>
<td>No</td>
<td>HDD *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>I3</td>
<td>No</td>
<td>No</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>I3en</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>M4</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>m4.16xlarge: ENA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All other sizes: Intel 82599 VF</td>
</tr>
<tr>
<td>M5</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>M5a</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>M5ad</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>M5d</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>P2</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>P3</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>P3dn</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>R4</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>R5</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>R5a</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>R5ad</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>R5d</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>T2</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>T3</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>ENA</td>
</tr>
<tr>
<td>T3a</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>ENA</td>
</tr>
<tr>
<td>u-xtb1.metal</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>ENA</td>
</tr>
<tr>
<td>X1</td>
<td>No</td>
<td>No</td>
<td>SSD</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>X1e</td>
<td>No</td>
<td>No</td>
<td>SSD *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
<tr>
<td>z1d</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
<td>Yes</td>
<td>ENA</td>
</tr>
</tbody>
</table>
* The root device volume must be an Amazon EBS volume.

The following table summarizes the networking and storage features supported by previous generation instance types.

<table>
<thead>
<tr>
<th>Instance store</th>
<th>Placement group</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3 SSD</td>
<td>Yes</td>
<td>Intel 82599 VF</td>
</tr>
<tr>
<td>G2 SSD</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>I2 SSD</td>
<td>Yes</td>
<td>Intel 82599 VF</td>
</tr>
<tr>
<td>M3 SSD</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>R3 SSD</td>
<td>Yes</td>
<td>Intel 82599 VF</td>
</tr>
</tbody>
</table>

**Instance Limits**

There is a limit on the total number of instances that you can launch in a region, and there are additional limits on some instance types.

For more information about the default limits, see How many instances can I run in Amazon EC2?

For more information about viewing your current limits or requesting an increase in your current limits, see Amazon EC2 Service Limits (p. 974).

**General Purpose Instances**

General purpose instances provide a balance of compute, memory, and networking resources, and can be used for a variety of workloads.

**M5, M5a, M5ad, M5d Instances**

These instances provide an ideal cloud infrastructure, offering a balance of compute, memory, and networking resources for a broad range of applications that are deployed in the cloud. M5 instances are well-suited for the following applications:

- Web and application servers
- Small and medium databases
- Gaming servers
- Caching fleets
- Running backend servers for SAP, Microsoft SharePoint, cluster computing, and other enterprise applications

m5.metal and m5d.metal instances provide your applications with direct access to physical resources of the host server, such as processors and memory. These instances are well suited for the following:

- Workloads that require access to low-level hardware features (for example, Intel VT) that are not available or fully supported in virtualized environments
- Applications that require a non-virtualized environment for licensing or support

For more information, see Amazon EC2 M5 Instances.
T2, T3, and T3a Instances

These instances provide a baseline level of CPU performance with the ability to burst to a higher level when required by your workload. An Unlimited instance can sustain high CPU performance for any period of time whenever required. For more information, see Burstable Performance Instances (p. 132). These instances are well-suited for the following applications:

- Websites and web applications
- Code repositories
- Development, build, test, and staging environments
- Microservices

For more information, see Amazon EC2 T2 Instances and Amazon EC2 T3 Instances.

Contents

- Hardware Specifications (p. 126)
- Instance Performance (p. 128)
- Network Performance (p. 128)
- SSD I/O Performance (p. 129)
- Instance Features (p. 130)
- Release Notes (p. 131)
- Burstable Performance Instances (p. 132)

Hardware Specifications

The following is a summary of the hardware specifications for general purpose instances.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Memory (GiB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m4.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>m4.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>m4.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>m4.4xlarge</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>m4.10xlarge</td>
<td>40</td>
<td>160</td>
</tr>
<tr>
<td>m4.16xlarge</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>m5.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>m5.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>m5.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>m5.4xlarge</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>m5.8xlarge</td>
<td>32</td>
<td>128</td>
</tr>
<tr>
<td>m5.12xlarge</td>
<td>48</td>
<td>192</td>
</tr>
<tr>
<td>m5.16xlarge</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>m5.24xlarge</td>
<td>96</td>
<td>384</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Memory (GiB)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>m5.metal</td>
<td>96</td>
<td>384</td>
</tr>
<tr>
<td>m5a.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>m5a.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>m5a.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>m5a.4xlarge</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>m5a.8xlarge</td>
<td>32</td>
<td>128</td>
</tr>
<tr>
<td>m5a.12xlarge</td>
<td>48</td>
<td>192</td>
</tr>
<tr>
<td>m5a.16xlarge</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>m5a.24xlarge</td>
<td>96</td>
<td>384</td>
</tr>
<tr>
<td>m5ad.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>m5ad.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>m5ad.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>m5ad.4xlarge</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>m5ad.12xlarge</td>
<td>48</td>
<td>192</td>
</tr>
<tr>
<td>m5ad.24xlarge</td>
<td>96</td>
<td>384</td>
</tr>
<tr>
<td>m5d.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>m5d.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>m5d.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>m5d.4xlarge</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>m5d.8xlarge</td>
<td>32</td>
<td>128</td>
</tr>
<tr>
<td>m5d.12xlarge</td>
<td>48</td>
<td>192</td>
</tr>
<tr>
<td>m5d.16xlarge</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>m5d.24xlarge</td>
<td>96</td>
<td>384</td>
</tr>
<tr>
<td>m5d.metal</td>
<td>96</td>
<td>384</td>
</tr>
<tr>
<td>t2.nano</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>t2.micro</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>t2.small</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>t2.medium</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>t2.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>t2.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>t2.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Memory (GiB)</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>t3.nano</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>t3.micro</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>t3.small</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t3.medium</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>t3.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>t3.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>t3.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>t3a.nano</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>t3a.micro</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>t3a.small</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t3a.medium</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>t3a.large</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>t3a.xlarge</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>t3a.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
</tbody>
</table>

For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.

For more information about specifying CPU options, see Optimizing CPU Options (p. 447).

**Instance Performance**

EBS-optimized instances enable you to get consistently high performance for your EBS volumes by eliminating contention between Amazon EBS I/O and other network traffic from your instance. Some general purpose instances are EBS-optimized by default at no additional cost. For more information, see Amazon EBS–Optimized Instances (p. 876).

**Network Performance**

You can enable enhanced networking capabilities on supported instance types. Enhanced networking provides significantly higher packet-per-second (PPS) performance, lower network jitter, and lower latencies. For more information, see Enhanced Networking on Windows (p. 700).

Instance types that use the Elastic Network Adapter (ENA) for enhanced networking deliver high packet per second performance with consistently low latencies. Most applications do not consistently need a high level of network performance, but can benefit from having access to increased bandwidth when they send or receive data. Instance sizes that use the ENA and are documented with network performance of "Up to 10 Gbps" or "Up to 25 Gbps" use a network I/O credit mechanism to allocate network bandwidth to instances based on average bandwidth utilization. These instances accrue credits when their network bandwidth is below their baseline limits, and can use these credits when they perform network data transfers.

The following is a summary of network performance for general purpose instances that support enhanced networking.
### Instance type | Network performance | Enhanced networking
---|---|---
`t2.nano | t2.micro | t2.small | t2.medium | t2.large | t2.xlarge | t2.2xlarge` | Up to 1 Gbps | 
`t3.nano | t3.micro | t3.small | t3.medium | t3.large | t3.xlarge | t3.2xlarge | t3a.nano | t3a.micro | t3a.small | t3a.medium | t3a.large | t3a.xlarge | t3a.2xlarge` | Up to 5 Gbps | ENA (p. 701)

m4.large | Moderate | Intel 82599 VF (p. 707)
m4.xlarge | m4.2xlarge | m4.4xlarge

m5.large | m5.xlarge | m5.2xlarge | m5.4xlarge | m5a.large | m5a.xlarge | m5a.2xlarge | m5a.4xlarge | m5a.8xlarge | m5ad.large | m5ad.xlarge | m5ad.2xlarge | m5ad.4xlarge | m5d.large | m5d.xlarge | m5d.2xlarge | m5d.4xlarge | Up to 10 Gbps | ENA (p. 701)

m4.10xlarge | 10 Gbps | Intel 82599 VF (p. 707)
m5.8xlarge | m5.12xlarge | m5a.12xlarge | m5ad.12xlarge | m5d.8xlarge | m5d.12xlarge

m5a.16xlarge | m5ad.16xlarge | 12 Gbps | ENA (p. 701)
m5.16xlarge | m5a.24xlarge | m5ad.24xlarge | m5d.16xlarge | 20 Gbps | ENA (p. 701)
m4.16xlarge | m5.24xlarge | m5.metal | m5d.24xlarge | m5d.metal | 25 Gbps | ENA (p. 701)

## SSD I/O Performance

If you use all the SSD-based instance store volumes available to your instance, you get the IOPS (4,096 byte block size) performance listed in the following table (at queue depth saturation). Otherwise, you get lower IOPS performance.
<table>
<thead>
<tr>
<th>Instance Size</th>
<th>100% Random Read IOPS</th>
<th>Write IOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>m5ad.large *</td>
<td>30,000</td>
<td>15,000</td>
</tr>
<tr>
<td>m5ad.xlarge *</td>
<td>59,000</td>
<td>29,000</td>
</tr>
<tr>
<td>m5ad.2xlarge *</td>
<td>117,000</td>
<td>57,000</td>
</tr>
<tr>
<td>m5ad.4xlarge *</td>
<td>234,000</td>
<td>114,000</td>
</tr>
<tr>
<td>m5ad.12xlarge</td>
<td>700,000</td>
<td>340,000</td>
</tr>
<tr>
<td>m5ad.24xlarge</td>
<td>1,400,000</td>
<td>680,000</td>
</tr>
<tr>
<td>m5d.large *</td>
<td>30,000</td>
<td>15,000</td>
</tr>
<tr>
<td>m5d.xlarge *</td>
<td>59,000</td>
<td>29,000</td>
</tr>
<tr>
<td>m5d.2xlarge *</td>
<td>117,000</td>
<td>57,000</td>
</tr>
<tr>
<td>m5d.4xlarge *</td>
<td>234,000</td>
<td>114,000</td>
</tr>
<tr>
<td>m5d.8xlarge</td>
<td>466,666</td>
<td>233,333</td>
</tr>
<tr>
<td>m5d.12xlarge</td>
<td>700,000</td>
<td>340,000</td>
</tr>
<tr>
<td>m5d.16xlarge</td>
<td>933,333</td>
<td>466,666</td>
</tr>
<tr>
<td>m5d.24xlarge</td>
<td>1,400,000</td>
<td>680,000</td>
</tr>
<tr>
<td>m5d.metal</td>
<td>1,400,000</td>
<td>680,000</td>
</tr>
</tbody>
</table>

* For these instances, you can get up to the specified performance.

As you fill the SSD-based instance store volumes for your instance, the number of write IOPS that you can achieve decreases. This is due to the extra work the SSD controller must do to find available space, rewrite existing data, and erase unused space so that it can be rewritten. This process of garbage collection results in internal write amplification to the SSD, expressed as the ratio of SSD write operations to user write operations. This decrease in performance is even larger if the write operations are not in multiples of 4,096 bytes or not aligned to a 4,096-byte boundary. If you write a smaller amount of bytes or bytes that are not aligned, the SSD controller must read the surrounding data and store the result in a new location. This pattern results in significantly increased write amplification, increased latency, and dramatically reduced I/O performance.

SSD controllers can use several strategies to reduce the impact of write amplification. One such strategy is to reserve space in the SSD instance storage so that the controller can more efficiently manage the space available for write operations. This is called over-provisioning. The SSD-based instance store volumes provided to an instance don’t have any space reserved for over-provisioning. To reduce write amplification, we recommend that you leave 10% of the volume unpartitioned so that the SSD controller can use it for over-provisioning. This decreases the storage that you can use, but increases performance even if the disk is close to full capacity.

For instance store volumes that support TRIM, you can use the TRIM command to notify the SSD controller whenever you no longer need data that you’ve written. This provides the controller with more free space, which can reduce write amplification and increase performance. For more information, see Instance Store Volume TRIM Support (p. 925).

**Instance Features**

The following is a summary of features for general purpose instances:
The root device volume must be an Amazon EBS volume.

For more information, see the following:

- Amazon EBS and NVMe on Windows Instances (p. 875)
- Amazon EC2 Instance Store (p. 917)
- Placement Groups (p. 710)

### Release Notes

- M5, M5d, and T3 instances feature a 3.1 GHz Intel Xeon Platinum 8000 series processor.
- M5a, M5ad, and T3a instances feature a 2.5 GHz AMD EPYC 7000 series processor.
- M4, M5, M5a, M5ad, M5d, t2.large and larger, and t3.large and larger, and t3a.large and larger instance types require 64-bit HVM AMIs. They have high-memory, and require a 64-bit operating system to take advantage of that capacity. HVM AMIs provide superior performance in comparison to paravirtual (PV) AMIs on high-memory instance types. In addition, you must use an HVM AMI to take advantage of enhanced networking.
- M5, M5a, M5ad, M5d, T3, and T3a instances have the following requirements:
  - NVMe drivers must be installed. EBS volumes are exposed as NVMe block devices (p. 875).
  - Elastic Network Adapter (ENA (p. 701)) drivers must be installed.

The following AMIs meet these requirements:

- Amazon Linux 2
- Amazon Linux AMI 2018.03
- Ubuntu 14.04 (with linux-aws kernel) or later
- Red Hat Enterprise Linux 7.4 or later
- SUSE Linux Enterprise Server 12 SP2 or later
- CentOS 7.4.1708 or later
- FreeBSD 11.1 or later
- M5, M5a, M5ad, M5d, T3, and T3a instances support a maximum of 28 attachments, including network interfaces, EBS volumes, and NVMe instance store volumes. Every instance has at least one network interface attachment. For example, if you have no additional network interface attachments on an EBS-only instance, you could attach 27 EBS volumes to that instance.
• Launching a bare metal instance boots the underlying server, which includes verifying all hardware and firmware components. This means that it can take 20 minutes from the time the instance enters the running state until it becomes available over the network.

• To attach or detach EBS volumes or secondary network interfaces from a bare metal instance requires PCIe native hotplug support.

• Bare metal instances use a PCI-based serial device rather than an I/O port-based serial device. The upstream Linux kernel and the latest Amazon Linux AMIs support this device. Bare metal instances also provide an ACPI SPCR table to enable the system to automatically use the PCI-based serial device. The latest Windows AMIs automatically use the PCI-based serial device.

• There is a limit on the total number of instances that you can launch in a region, and there are additional limits on some instance types. For more information, see How many instances can I run in Amazon EC2?. To request a limit increase, use the Amazon EC2 Instance Request Form.

Burstable Performance Instances

Burstable performance instances, which are T3, T3a, and T2 instances, are designed to provide a baseline level of CPU performance with the ability to burst to a higher level when required by your workload. Burstable performance instances are well suited for a wide range of general-purpose applications. Examples include microservices, low-latency interactive applications, small and medium databases, virtual desktops, development, build, and stage environments, code repositories, and product prototypes.

Burstable performance instances are the only instance types that use credits for CPU usage. For more information about instance pricing and additional hardware details, see Amazon EC2 Pricing and Amazon EC2 Instance Types.

If your account is less than 12 months old, you can use a `t2.micro` instance for free within certain usage limits. For more information, see AWS Free Tier.

Contents

• Burstable Performance Instance Requirements (p. 132)
• Best Practices (p. 133)
• CPU Credits and Baseline Performance for Burstable Performance Instances (p. 133)
• Unlimited Mode for Burstable Performance Instances (p. 135)
• Standard Mode for Burstable Performance Instances (p. 142)
• Working with Burstable Performance Instances (p. 153)
• Monitoring Your CPU Credits (p. 157)

Burstable Performance Instance Requirements

The following are the requirements for these instances:

• These instances are available as On-Demand Instances, Reserved Instances, and Spot Instances, but not as Scheduled Instances or Dedicated Instances. They are also not supported on a Dedicated Host. For more information, see Instance Purchasing Options (p. 191).

• Ensure that the instance size you choose passes the minimum memory requirements of your operating system and applications. Operating systems with graphical user interfaces that consume significant memory and CPU resources (for example, Windows) might require a `t2.micro` or larger instance size for many use cases. As the memory and CPU requirements of your workload grow over time, you can scale to larger instance sizes of the same instance type, or another instance type.

• For additional requirements, see General Purpose Instances Release Notes (p. 131).
Best Practices

Follow these best practices to get the maximum benefit from burstable performance instances.

- **Use a recommended AMI** – Use an AMI that provides the required drivers. For more information, see Release Notes (p. 131).
- **Turn on instance recovery** – Create a CloudWatch alarm that monitors an EC2 instance and automatically recovers it if it becomes impaired for any reason. For more information, see Adding Recover Actions to Amazon CloudWatch Alarms (p. 569).

CPU Credits and Baseline Performance for Burstable Performance Instances

Traditional Amazon EC2 instance types provide fixed performance, while burstable performance instances provide a baseline level of CPU performance with the ability to burst above that baseline level. The baseline performance and ability to burst are governed by CPU credits. A CPU credit provides the performance of a full CPU core for one minute.

### Contents

- CPU Credits (p. 133)
- Baseline Performance (p. 135)

#### CPU Credits

One CPU credit is equal to one vCPU running at 100% utilization for one minute. Other combinations of number of vCPUs, utilization, and time can also equate to one CPU credit. For example, one CPU credit is equal to one vCPU running at 50% utilization for two minutes, or two vCPUs running at 25% utilization for two minutes.

#### Earning CPU Credits

Each burstable performance instance continuously earns (at a millisecond-level resolution) a set rate of CPU credits per hour, depending on the instance size. The accounting process for whether credits are accrued or spent also happens at a millisecond-level resolution, so you don't have to worry about overspending CPU credits; a short burst of CPU uses a small fraction of a CPU credit.

If a burstable performance instance uses fewer CPU resources than is required for baseline performance (such as when it is idle), the unspent CPU credits are accrued in the CPU credit balance. If a burstable performance instance needs to burst above the baseline performance level, it spends the accrued credits. The more credits that a burstable performance instance has accrued, the more time it can burst beyond its baseline when more performance is needed.

The following table lists the burstable performance instance types, the rate at which CPU credits are earned per hour, the maximum number of earned CPU credits that an instance can accrue, the number of vCPUs per instance, and the baseline performance level as a percentage of a full core performance (using a single vCPU).

<table>
<thead>
<tr>
<th>Instance type</th>
<th>CPU credits earned per hour</th>
<th>Maximum earned credits that can be accrued*</th>
<th>vCPUs</th>
<th>Baseline performance per vCPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>t2.nano</td>
<td>3</td>
<td>72</td>
<td>1</td>
<td>5%</td>
</tr>
<tr>
<td>t2.micro</td>
<td>6</td>
<td>144</td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>t2.small</td>
<td>12</td>
<td>288</td>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td>t2.medium</td>
<td>24</td>
<td>576</td>
<td>2</td>
<td>20%**</td>
</tr>
<tr>
<td>Instance type</td>
<td>CPU credits earned per hour</td>
<td>Maximum earned credits that can be accrued*</td>
<td>vCPUs</td>
<td>Baseline performance per vCPU</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------</td>
<td>---------------------------------------------</td>
<td>-------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>t2.large</td>
<td>36</td>
<td>864</td>
<td>2</td>
<td>30%**</td>
</tr>
<tr>
<td>t2.xlarge</td>
<td>54</td>
<td>1296</td>
<td>4</td>
<td>22.5%**</td>
</tr>
<tr>
<td>t2.2xlarge</td>
<td>81.6</td>
<td>1958.4</td>
<td>8</td>
<td>17%**</td>
</tr>
<tr>
<td>t3.nano</td>
<td>6</td>
<td>144</td>
<td>2</td>
<td>5%**</td>
</tr>
<tr>
<td>t3.micro</td>
<td>12</td>
<td>288</td>
<td>2</td>
<td>10%**</td>
</tr>
<tr>
<td>t3.small</td>
<td>24</td>
<td>576</td>
<td>2</td>
<td>20%**</td>
</tr>
<tr>
<td>t3.medium</td>
<td>24</td>
<td>576</td>
<td>2</td>
<td>20%**</td>
</tr>
<tr>
<td>t3.large</td>
<td>36</td>
<td>864</td>
<td>2</td>
<td>30%**</td>
</tr>
<tr>
<td>t3.xlarge</td>
<td>96</td>
<td>2304</td>
<td>4</td>
<td>40%**</td>
</tr>
<tr>
<td>t3.2xlarge</td>
<td>192</td>
<td>4608</td>
<td>8</td>
<td>40%**</td>
</tr>
<tr>
<td>t3a.nano</td>
<td>6</td>
<td>144</td>
<td>2</td>
<td>5%**</td>
</tr>
<tr>
<td>t3a.micro</td>
<td>12</td>
<td>288</td>
<td>2</td>
<td>10%**</td>
</tr>
<tr>
<td>t3a.small</td>
<td>24</td>
<td>576</td>
<td>2</td>
<td>20%**</td>
</tr>
<tr>
<td>t3a.medium</td>
<td>24</td>
<td>576</td>
<td>2</td>
<td>20%**</td>
</tr>
<tr>
<td>t3a.large</td>
<td>36</td>
<td>864</td>
<td>2</td>
<td>30%**</td>
</tr>
<tr>
<td>t3a.xlarge</td>
<td>96</td>
<td>2304</td>
<td>4</td>
<td>40%**</td>
</tr>
<tr>
<td>t3a.2xlarge</td>
<td>192</td>
<td>4608</td>
<td>8</td>
<td>40%**</td>
</tr>
</tbody>
</table>

* The number of credits that can be accrued is equivalent to the number of credits that can be earned in a 24-hour period.

** The baseline performance in the table is per vCPU. For instance sizes that have more than one vCPU, to calculate the baseline CPU utilization for the instance, multiply the vCPU percentage by the number of vCPUs. For example, a t3.large instance has two vCPUs, which provide a baseline CPU utilization for the instance of 60% (2 vCPUs x 30% baseline performance of one vCPU). In CloudWatch, CPU utilization is shown per vCPU. Therefore, the CPU utilization for a t3.large instance operating at the baseline performance is shown as 30% in CloudWatch CPU metrics.

**CPU Credit Earn Rate**

The number of CPU credits earned per hour is determined by the instance size. For example, a t3.nano earns six credits per hour, while a t3.small earns 24 credits per hour. The preceding table lists the credit earn rate for all instances.

**CPU Credit Accrual Limit**

While earned credits never expire on a running instance, there is a limit to the number of earned credits that an instance can accrue. The limit is determined by the CPU credit balance limit. After the limit is
reached, any new credits that are earned are discarded, as indicated by the following image. The full bucket indicates the CPU credit balance limit, and the spillover indicates the newly earned credits that exceed the limit.

![CPU Credit Balance Diagram](image)

The CPU credit balance limit differs for each instance size. For example, a t3.micro instance can accrue a maximum of 288 earned CPU credits in the CPU credit balance. The preceding table lists the maximum number of earned credits that each instance can accrue.

**Note**

T2 Standard instances also earn launch credits. Launch credits do not count towards the CPU credit balance limit. If a T2 instance has not spent its launch credits, and remains idle over a 24-hour period while accruing earned credits, its CPU credit balance appears as over the limit. For more information, see Launch Credits (p. 143). T3 and T3a instances do not earn launch credits. These instances launch as unlimited by default, and therefore can burst immediately upon start without any launch credits.

**Accrued CPU Credits Life Span**

CPU credits on a running instance do not expire.

For T3 and T3a, the CPU credit balance persists for seven days after an instance stops and the credits are lost thereafter. If you start the instance within seven days, no credits are lost.

For T2, the CPU credit balance does not persist between instance stops and starts. If you stop a T2 instance, the instance loses all its accrued credits.

For more information, see CPUCreditBalance in the CloudWatch metrics table (p. 157).

**Baseline Performance**

The number of credits that an instance earns per hour can be expressed as a percentage of CPU utilization. It is known as the baseline performance, and sometimes just as the baseline. For example, a t3.nano instance, with two vCPUs, earns six credits per hour, resulting in a baseline performance of 5% (3/60 minutes) per vCPU. A t3.xlarge instance, with four vCPUs, earns 96 credits per hour, resulting in a baseline performance of 40% (24/60 minutes) per vCPU.

**Unlimited Mode for Burstable Performance Instances**

A burstable performance instance configured as unlimited can sustain high CPU performance for any period of time whenever required. The hourly instance price automatically covers all CPU usage spikes if the average CPU utilization of the instance is at or below the baseline over a rolling 24-hour period or the instance lifetime, whichever is shorter.

For the vast majority of general-purpose workloads, instances configured as unlimited provide ample performance without any additional charges. If the instance runs at higher CPU utilization for a prolonged period, it can do so for a flat additional rate per vCPU-hour. For information about instance pricing, see Amazon EC2 Pricing and the section for Unlimited pricing in Amazon EC2 On-Demand Pricing.
Important
If you use a t2.micro instance under the AWS Free Tier offer and configure it as unlimited, charges may apply if your average utilization over a rolling 24-hour period exceeds the baseline of the instance.

Contents
• Unlimited Mode Concepts (p. 136)
• Examples: Unlimited Mode (p. 139)

Unlimited Mode Concepts
The unlimited mode is a credit configuration option for burstable performance instances. It can be enabled or disabled at any time for a running or stopped instance.

Note
T3 and T3a instances are launched as unlimited by default. T2 instances are launched as standard by default.

How Unlimited Burstable Performance Instances Work
If a burstable performance instance configured as unlimited depletes its CPU credit balance, it can spend surplus credits to burst beyond the baseline. When its CPU utilization falls below the baseline, it uses the CPU credits that it earns to pay down the surplus credits that it spent earlier. The ability to earn CPU credits to pay down surplus credits enables Amazon EC2 to average the CPU utilization of an instance over a 24-hour period. If the average CPU usage over a 24-hour period exceeds the baseline, the instance is billed for the additional usage at a flat additional rate per vCPU-hour.

The following graph shows the CPU usage of a t3.large. The baseline CPU utilization for a t3.large is 30%. If the instance runs at 30% CPU utilization or less on average over a 24-hour period, there is no additional charge because the cost is already covered by the instance hourly price. However, if the instance runs at 40% CPU utilization on average over a 24-hour period, as shown in the graph, the instance is billed for the additional 10% CPU usage at a flat additional rate per vCPU-hour.

For more information about the baseline performance per vCPU for each instance type and how many credits each instance type earns, see the credit table (p. 133).

When to Use Unlimited Mode vs Fixed CPU
When determining whether you should use a burstable performance instance in unlimited mode, such as a T3, or a fixed performance instance, such as an M5, you need to determine the breakeven CPU usage. The breakeven CPU usage for a burstable performance instance is the point at which a burstable performance instance costs the same as a fixed performance instance. The breakeven CPU usage helps you determine the following:
• If the average CPU usage over a 24-hour period is at or below the breakeven CPU usage, use a burstable performance instance in unlimited mode so that you can benefit from the lower price of a burstable performance instance while getting the same performance as a fixed performance instance.

• If the average CPU usage over a 24-hour period is above the breakeven CPU usage, the burstable performance instance will cost more than the equivalently-sized fixed performance instance. If a T3 instance continuously bursts at 100% CPU, you end up paying approximately 1.5 times the price of an equivalently-sized M5 instance.

The following graph shows the breakeven CPU usage point where a t3.large costs the same as an m5.large. The breakeven CPU usage point for a t3.large is 42.5%. If the average CPU usage is at 42.5%, the cost of running the t3.large is the same as an m5.large, and is more expensive if the average CPU usage is above 42.5%. If the workload needs less than 42.5% average CPU usage, you can benefit from the lower price of the t3.large while getting the same performance as an m5.large.

The following table shows how to calculate the breakeven CPU usage threshold so that you can determine when it's less expensive to use a burstable performance instance in unlimited mode or a fixed performance instance. The columns in the table are labeled A through K.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>vCPUs</th>
<th>T3 price*/hour</th>
<th>M5 price*/hour</th>
<th>Price difference</th>
<th>Charge per vCPU</th>
<th>Charge per vCPU minute</th>
<th>Charge additional burst minutes available</th>
<th>Additional CPU %</th>
<th>Breakeven CPU %</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3.large</td>
<td>2</td>
<td>$0.0835</td>
<td>$0.096</td>
<td>$0.0125</td>
<td>$0.05</td>
<td>$0.000833</td>
<td>15</td>
<td>12.5%</td>
<td>42.5%</td>
</tr>
</tbody>
</table>

* Price is based on us-east-1 and Linux OS.
• Column A shows the instance type, t3.large.
• Column B shows the number of vCPUs for the t3.large.
• Column C shows the price of a t3.large per hour.
• Column D shows the price of an m5.large per hour.
• Column E shows the price difference between the t3.large and the m5.large.
• Column F shows the baseline performance per vCPU of the t3.large, which is 30%. At the baseline, the hourly cost of the instance covers the cost of the CPU usage.
• Column G shows the flat additional rate per vCPU-hour that an instance is charged if it bursts at 100% CPU after it has depleted its earned credits.
• Column H shows the flat additional rate per vCPU-minute that an instance is charged if it bursts at 100% CPU after it has depleted its earned credits.
• Column I shows the number of additional minutes that the t3.large can burst per hour at 100% CPU while paying the same price per hour as an m5.large.
• Column J shows the additional CPU usage (in %) over baseline that the instance can burst while paying the same price per hour as an m5.large.
• Column K shows the breakeven CPU usage (in %) that the t3.large can burst without paying more than the m5.large. Anything above this, and the t3.large costs more than the m5.large.

The following table shows the breakeven CPU usage (in %) for T3 instance types compared to the similarly-sized M5 instance types.

<table>
<thead>
<tr>
<th>T3 instance type</th>
<th>Breakeven CPU usage (in %) for T3 compared to M5</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3.large</td>
<td>42.5%</td>
</tr>
<tr>
<td>t3.xlarge</td>
<td>52.5%</td>
</tr>
<tr>
<td>t3.2xlarge</td>
<td>52.5%</td>
</tr>
</tbody>
</table>

**Surplus Credits Can Incur Charges**

If the average CPU utilization of an instance is at or below the baseline, the instance incurs no additional charges. Because an instance earns a maximum number of credits (p. 133) in a 24-hour period (for example, a t3.micro instance can earn a maximum of 288 credits in a 24-hour period), it can spend surplus credits up to that maximum without being charged.

However, if CPU utilization stays above the baseline, the instance cannot earn enough credits to pay down the surplus credits that it has spent. The surplus credits that are not paid down are charged at a flat additional rate per vCPU-hour.

Surplus credits that were spent earlier are charged when any of the following occurs:

• The spent surplus credits exceed the maximum number of credits (p. 133) the instance can earn in a 24-hour period. Spent surplus credits above the maximum are charged at the end of the hour.
• The instance is stopped or terminated.
• The instance is switched from unlimited to standard.

Spent surplus credits are tracked by the CloudWatch metric CPUSurplusCreditBalance. Surplus credits that are charged are tracked by the CloudWatch metric CPUSurplusCreditsCharged. For more information, see Additional CloudWatch Metrics for Burstable Performance Instances (p. 157).
No Launch Credits for T2 Unlimited

T2 Standard instances receive launch credits (p. 143), but T2 Unlimited instances do not. A T2 Unlimited instance can burst beyond the baseline at any time with no additional charge, as long as its average CPU utilization is at or below the baseline over a rolling 24-hour window or its lifetime, whichever is shorter. As such, T2 Unlimited instances do not require launch credits to achieve high performance immediately after launch.

If a T2 instance is switched from standard to unlimited, any accrued launch credits are removed from the CPUCreditBalance before the remaining CPUCreditBalance is carried over.

Note

T3 and T3a instances never receive launch credits.

Enabling Unlimited Mode

T3 and T3a instances launch as unlimited by default. T2 instances launch as standard by default, but you can enable unlimited at launch.

You can switch from unlimited to standard, and from standard to unlimited, at any time on a running or stopped instance. For more information, see Launching a Burstable Performance Instance as Unlimited or Standard (p. 154) and Modifying the Credit Specification of a Burstable Performance Instance (p. 156).

You can check whether your burstable performance instance is configured as unlimited or standard using the Amazon EC2 console or the AWS CLI. For more information, see Viewing the Credit Specification of a Burstable Performance Instance (p. 156).

What Happens to Credits when Switching between Unlimited and Standard

CPUCreditBalance is a CloudWatch metric that tracks the number of credits accrued by an instance. CPUSurplusCreditBalance is a CloudWatch metric that tracks the number of surplus credits spent by an instance.

When you change an instance configured as unlimited to standard, the following occurs:

- The CPUCreditBalance value remains unchanged and is carried over.
- The CPUSurplusCreditBalance value is immediately charged.

When a standard instance is switched to unlimited, the following occurs:

- The CPUCreditBalance value containing accrued earned credits is carried over.
- For T2 Standard instances, any launch credits are removed from the CPUCreditBalance value, and the remaining CPUCreditBalance value containing accrued earned credits is carried over.

Monitoring Credit Usage

To see if your instance is spending more credits than the baseline provides, you can use CloudWatch metrics to track usage, and you can set up hourly alarms to be notified of credit usage. For more information, see Monitoring Your CPU Credits (p. 157).

Examples: Unlimited Mode

The following examples explain credit use for instances that are configured as unlimited.

Examples
Example 1: Explaining Credit Use with T3 Unlimited

In this example, you see the CPU utilization of a t3.nano instance launched as unlimited, and how it spends earned and surplus credits to sustain CPU performance.

A t3.nano instance earns 144 CPU credits over a rolling 24-hour period, which it can redeem for 144 minutes of vCPU use. When it depletes its CPU credit balance (represented by the CloudWatch metric CPUCreditBalance), it can spend surplus CPU credits—that it has not yet earned—to burst for as long as it needs. Because a t3.nano instance earns a maximum of 144 credits in a 24-hour period, it can spend surplus credits up to that maximum without being charged immediately. If it spends more than 144 CPU credits, it is charged for the difference at the end of the hour.

The intent of the example, illustrated by the following graph, is to show how an instance can burst using surplus credits even after it depletes its CPUCreditBalance. The following workflow references the numbered points on the graph:

P1 – At 0 hours on the graph, the instance is launched as unlimited and immediately begins to earn credits. The instance remains idle from the time it is launched—CPU utilization is 0%—and no credits are spent. All unspent credits are accrued in the credit balance. For the first 24 hours, CPUCreditUsage is at 0, and the CPUCreditBalance value reaches its maximum of 144.

P2 – For the next 12 hours, CPU utilization is at 2.5%, which is below the 5% baseline. The instance earns more credits than it spends, but the CPUCreditBalance value cannot exceed its maximum of 144 credits.

P3 – For the next 24 hours, CPU utilization is at 7% (above the baseline), which requires a spend of 57.6 credits. The instance spends more credits than it earns, and the CPUCreditBalance value reduces to 86.4 credits.

P4 – For the next 12 hours, CPU utilization decreases to 2.5% (below the baseline), which requires a spend of 36 credits. In the same time, the instance earns 72 credits. The instance earns more credits than it spends, and the CPUCreditBalance value increases to 122 credits.

P5 – For the next 5 hours, the instance bursts at 100% CPU utilization, and spends a total of 570 credits to sustain the burst. About an hour into this period, the instance depletes its entire CPUCreditBalance of 122 credits, and starts to spend surplus credits to sustain the high CPU performance, totaling 448 surplus credits in this period (570-122=448). When the CPUSurplusCreditBalance value reaches 144 CPU credits (the maximum a t3.nano instance can earn in a 24-hour period), any surplus credits spent thereafter cannot be offset by earned credits. The surplus credits spent thereafter amounts to 304 credits (448-144=304), which results in a small additional charge at the end of the hour for 304 credits.

P6 – For the next 13 hours, CPU utilization is at 5% (the baseline). The instance earns as many credits as it spends, with no excess to pay down the CPUSurplusCreditBalance. The CPUSurplusCreditBalance value remains at 144 credits.

P7 – For the last 24 hours in this example, the instance is idle and CPU utilization is 0%. During this time, the instance earns 144 credits, which it uses to pay down the CPUSurplusCreditBalance.
Example 2: Explaining Credit Use with T2 Unlimited

In this example, you see the CPU utilization of a t2.nano instance launched as unlimited, and how it spends eared and surplus credits to sustain CPU performance.

A t2.nano instance earns 72 CPU credits over a rolling 24-hour period, which it can redeem for 72 minutes of vCPU use. When it depletes its CPU credit balance (represented by the CloudWatch metric CPUCreditBalance), it can spend surplus CPU credits—that it has not yet earned—to burst for as long as it needs. Because a t2.nano instance earns a maximum of 72 credits in a 24-hour period, it can spend surplus credits up to that maximum without being charged immediately. If it spends more than 72 CPU credits, it is charged for the difference at the end of the hour.

The intent of the example, illustrated by the following graph, is to show how an instance can burst using surplus credits even after it depletes its CPUCreditBalance. You can assume that, at the start of the time line in the graph, the instance has an accrued credit balance equal to the maximum number of credits it can earn in 24 hours. The following workflow references the numbered points on the graph:

1 – In the first 10 minutes, CPUCreditUsage is at 0, and the CPUCreditBalance value remains at its maximum of 72.

2 – At 23:40, as CPU utilization increases, the instance spends CPU credits and the CPUCreditBalance value decreases.

3 – At around 00:47, the instance depletes its entire CPUCreditBalance, and starts to spend surplus credits to sustain high CPU performance.

4 – Surplus credits are spent until 01:55, when the CPUSurplusCreditBalance value reaches 72 CPU credits. This is equal to the maximum a t2.nano instance can earn in a 24-hour period. Any surplus
credits spent thereafter cannot be offset by earned credits within the 24-hour period, which results in a small additional charge at the end of the hour.

5. The instance continues to spend surplus credits until around 02:20. At this time, CPU utilization falls below the baseline, and the instance starts to earn credits at 3 credits per hour (or 0.25 credits every 5 minutes), which it uses to pay down the CPUSurplusCreditBalance. After the CPUSurplusCreditBalance value reduces to 0, the instance starts to accrue earned credits in its CPUCreditBalance at 0.25 credits every 5 minutes.

Calculating the Bill

Surplus credits cost $0.096 per vCPU-hour. The instance spent approximately 25 surplus credits between 01:55 and 02:20, which is equivalent to 0.42 vCPU-hours.

Additional charges for this instance are 0.42 vCPU-hours x $0.096/vCPU-hour = $0.04032, rounded to $0.04.

Here is the month-end bill for this T2 Unlimited instance:

<table>
<thead>
<tr>
<th>Amazon Elastic Compute Cloud running Windows</th>
<th>Amazon Elastic Compute Cloud T2 CPU Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.018 per On Demand Windows T2.m2-instance-hour</td>
<td>$0.04 per vCPU-hour of T2 CPU credits</td>
</tr>
<tr>
<td>$220.00/hr</td>
<td>9,456 vCPU-hours</td>
</tr>
<tr>
<td>$5.43</td>
<td>$0.04</td>
</tr>
</tbody>
</table>

You can set billing alerts to be notified every hour of any accruing charges, and take action if required.

Standard Mode for Burstable Performance Instances

A burstable performance instance configured as standard is suited to workloads with an average CPU utilization that is consistently below the baseline performance of the instance. To burst above the baseline, the instance spends credits that it has accrued in its CPU credit balance. If the instance is
running low on accrued credits, performance is gradually lowered to the baseline performance level, so that the instance does not experience a sharp performance drop-off when its accrued CPU credit balance is depleted. For more information, see CPU Credits and Baseline Performance for Burstable Performance Instances (p. 133).

Contents

- Standard Mode Concepts (p. 143)
- Examples: Standard Mode (p. 145)

Standard Mode Concepts

The standard mode is a configuration option for burstable performance instances. It can be enabled or disabled at any time for a running or stopped instance.

Note

T3 and T3a instances are launched as unlimited by default. T2 instances are launched as standard by default.

How Standard Burstable Performance Instances Work

When a burstable performance instance configured as standard is in a running state, it continuously earns (at a millisecond-level resolution) a set rate of earned credits per hour. For T2 Standard, when the instance is stopped, it loses all its accrued credits, and its credit balance is reset to zero. When it is restarted, it receives a new set of launch credits, and begins to accrue earned credits. For T3 and T3a Standard, the CPU credit balance persists for seven days after the instance stops and the credits are lost thereafter. If you start the instance within seven days, no credits are lost.

A T2 Standard instance receives two types of CPU credits: earned credits and launch credits. When a T2 Standard instance is in a running state, it continuously earns (at a millisecond-level resolution) a set rate of earned credits per hour. At start, it has not yet earned credits for a good startup experience; therefore, to provide a good startup experience, it receives launch credits at start, which it spends first while it accrues earned credits.

T3 and T3a Standard instances do not receive launch credits.

Launch Credits

T2 Standard instances get 30 launch credits per vCPU at launch or start. For example, a t2.micro instance has one vCPU and gets 30 launch credits, while a t2.xlarge instance has four vCPUs and gets 120 launch credits. Launch credits are designed to provide a good startup experience to allow instances to burst immediately after launch before they have accrued earned credits.

Launch credits are spent first, before earned credits. Unspent launch credits are accrued in the CPU credit balance, but do not count towards the CPU credit balance limit. For example, a t2.micro instance has a CPU credit balance limit of 144 earned credits. If it is launched and remains idle for 24 hours, its CPU credit balance reaches 174 (30 launch credits + 144 earned credits), which is over the limit. However, after the instance spends the 30 launch credits, the credit balance cannot exceed 144. For more information about the CPU credit balance limit for each instance size, see the credit table (p. 133).

The following table lists the initial CPU credit allocation received at launch or start, and the number of vCPUs.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Launch credits</th>
<th>vCPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1.micro</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>t2.nano</td>
<td>30</td>
<td>1</td>
</tr>
</tbody>
</table>
### Instance type | Launch credits | vCPUs
--- | --- | ---
t2.micro | 30 | 1
t2.small | 30 | 1
t2.medium | 60 | 2
t2.large | 60 | 2
t2.xlarge | 120 | 4
t2.2xlarge | 240 | 8

### Launch Credit Limits

There is a limit to the number of times T2 Standard instances can receive launch credits. The default limit is 100 launches or starts of all T2 Standard instances combined per account, per Region, per rolling 24-hour period. For example, the limit is reached when one instance is stopped and started 100 times within a 24-hour period, or when 100 instances are launched within a 24-hour period, or other combinations that equate to 100 starts. New accounts may have a lower limit, which increases over time based on your usage.

**Tip**

To ensure that your workloads always get the performance they need, switch to Unlimited Mode for Burstable Performance Instances (p. 135) or consider using a larger instance size.

### Differences Between Launch Credits and Earned Credits

The following table lists the differences between launch credits and earned credits.

<table>
<thead>
<tr>
<th>Credit earn rate</th>
<th>Launch credits</th>
<th>Earned credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2 Standard instances get 30 launch credits per vCPU at launch or start. If a T2 instance is switched from unlimited to standard, it does not get launch credits at the time of switching.</td>
<td>Each T2 instance continuously earns (at a millisecond-level resolution) a set rate of CPU credits per hour, depending on the instance size. For more information about the number of CPU credits earned per instance size, see the credit table (p. 133).</td>
<td></td>
</tr>
</tbody>
</table>

| Credit earn limit | The limit for receiving launch credits is 100 launches or starts of all T2 Standard instances combined per account, per Region, per rolling 24-hour period. New accounts may have a lower limit, which increases over time based on your usage. | A T2 instance cannot accrue more credits than the CPU credit balance limit. If the CPU credit balance has reached its limit, any credits that are earned after the limit is reached are discarded. Launch credits do not count towards the limit. For more information about the CPU credit balance limit for each T2 instance size, see the credit table (p. 133). |

| Credit use | Launch credits are spent first, before earned credits. | Earned credits are spent only after all launch credits are spent. |
| Credit expiration | When a T2 Standard instance is running, launch credits do not expire. When a T2 Standard instance stops or is switched to T2 Unlimited, all launch credits are lost. | When a T2 instance is running, earned credits that have accrued do not expire. When the T2 instance stops, all accrued earned credits are lost. |
The number of accrued launch credits and accrued earned credits is tracked by the CloudWatch metric CPUCreditBalance. For more information, see CPUCreditBalance in the CloudWatch metrics table (p. 157).

**Examples: Standard Mode**

The following examples explain credit use when instances are configured as standard.

**Examples**

- Example 1: Explaining Credit Use with T3 Standard (p. 145)
- Example 2: Explaining Credit Use with T2 Standard (p. 146)

**Example 1: Explaining Credit Use with T3 Standard**

In this example, you see how a t3.nano instance launched as standard earns, accrues, and spends earned credits. You see how the credit balance reflects the accrued earned credits.

**Note**

T3 and T3a instances configured as standard do not receive launch credits.

A running t3.nano instance earns 144 credits every 24 hours. Its credit balance limit is 144 earned credits. After the limit is reached, new credits that are earned are discarded. For more information about the number of credits that can be earned and accrued, see the credit table (p. 133).

You might launch a T3 Standard instance and use it immediately. Or, you might launch a T3 Standard instance and leave it idle for a few days before running applications on it. Whether an instance is used or remains idle determines if credits are spent or accrued. If an instance remains idle for 24 hours from the time it is launched, the credit balance reaches its limit, which is the maximum number of earned credits that can be accrued.

This example describes an instance that remains idle for 24 hours from the time it is launched, and walks you through seven periods of time over a 96-hour period, showing the rate at which credits are earned, accrued, spent, and discarded, and the value of the credit balance at the end of each period.

The following workflow references the numbered points on the graph:

**P1** – At 0 hours on the graph, the instance is launched as standard and immediately begins to earn credits. The instance remains idle from the time it is launched—CPU utilization is 0%—and no credits are spent. All unspent credits are accrued in the credit balance. For the first 24 hours, CPUCreditUsage is at 0, and the CPUCreditBalance value reaches its maximum of 144.

**P2** – For the next 12 hours, CPU utilization is at 2.5%, which is below the 5% baseline. The instance earns more credits than it spends, but the CPUCreditBalance value cannot exceed its maximum of 144 credits. Any credits that are earned in excess of the limit are discarded.

**P3** – For the next 24 hours, CPU utilization is at 7% (above the baseline), which requires a spend of 57.6 credits. The instance spends more credits than it earns, and the CPUCreditBalance value reduces to 86.4 credits.

**P4** – For the next 12 hours, CPU utilization decreases to 2.5% (below the baseline), which requires a spend of 36 credits. In the same time, the instance earns 72 credits. The instance earns more credits than it spends, and the CPUCreditBalance value increases to 122 credits.

**P5** – For the next two hours, the instance bursts at 100% CPU utilization, and depletes its entire CPUCreditBalance value of 122 credits. At the end of this period, with the CPUCreditBalance at zero, CPU utilization is forced to drop to the baseline performance level of 5%. At the baseline, the instance earns as many credits as it spends.
P6 – For the next 14 hours, CPU utilization is at 5% (the baseline). The instance earns as many credits as it spends. The CPUCreditBalance value remains at 0.

P7 – For the last 24 hours in this example, the instance is idle and CPU utilization is 0%. During this time, the instance earns 144 credits, which it accrues in its CPUCreditBalance.

Example 2: Explaining Credit Use with T2 Standard

In this example, you see how a t2.nano instance launched as standard earns, accrues, and spends launch and earned credits. You see how the credit balance reflects not only accrued earned credits, but also accrued launch credits.

A t2.nano instance gets 30 launch credits when it is launched, and earns 72 credits every 24 hours. Its credit balance limit is 72 earned credits; launch credits do not count towards the limit. After the limit is reached, new credits that are earned are discarded. For more information about the number of credits that can be earned and accrued, see the credit table (p. 133). For more information about limits, see Launch Credit Limits (p. 144).

You might launch a T2 Standard instance and use it immediately. Or, you might launch a T2 Standard instance and leave it idle for a few days before running applications on it. Whether an instance is used or remains idle determines if credits are spent or accrued. If an instance remains idle for 24 hours from the time it is launched, the credit balance appears to exceed its limit because the balance reflects both accrued earned credits and accrued launch credits. However, after CPU is used, the launch credits are spent first. Thereafter, the limit always reflects the maximum number of earned credits that can be accrued.

This example describes an instance that remains idle for 24 hours from the time it is launched, and walks you through seven periods of time over a 96-hour period, showing the rate at which credits are earned, accrued, spent, and discarded, and the value of the credit balance at the end of each period.

Period 1: 1 – 24 hours

At 0 hours on the graph, the T2 instance is launched as standard and immediately gets 30 launch credits. It earns credits while in the running state. The instance remains idle from the time it is launched—CPU utilization is 0%—and no credits are spent. All unspent credits are accrued in the credit balance. At approximately 14 hours after launch, the credit balance is 72 (30 launch credits + 42 earned credits), which is equivalent to what the instance can earn in 24 hours. At 24 hours after launch, the credit balance exceeds 72 credits because the unspent launch credits are accrued in the credit balance—the credit balance is 102 credits: 30 launch credits + 72 earned credits.
Credit Spend Rate
0 credits per 24 hours (0% CPU utilization)

Credit Earn Rate
72 credits per 24 hours

Credit Discard Rate
0 credits per 24 hours

Credit Balance
102 credits (30 launch credits + 72 earned credits)

Conclusion

If there is no CPU utilization after launch, the instance accrues more credits than what it can earn in 24 hours (30 launch credits + 72 earned credits = 102 credits).

In a real-world scenario, an EC2 instance consumes a small number of credits while launching and running, which prevents the balance from reaching the maximum theoretical value in this example.

Period 2: 25 – 36 hours

For the next 12 hours, the instance continues to remain idle and earn credits, but the credit balance does not increase. It plateaus at 102 credits (30 launch credits + 72 earned credits). The credit balance has reached its limit of 72 accrued earned credits, so newly earned credits are discarded.
Credit Spend Rate

0 credits per 24 hours (0% CPU utilization)

Credit Earn Rate

72 credits per 24 hours (3 credits per hour)

Credit Discard Rate

72 credits per 24 hours (100% of credit earn rate)

Credit Balance

102 credits (30 launch credits + 72 earned credits) — balance is unchanged

Conclusion

An instance constantly earns credits, but it cannot accrue more earned credits if the credit balance has reached its limit. After the limit is reached, newly earned credits are discarded. Launch credits do not count towards the credit balance limit. If the balance includes accrued launch credits, the balance appears to be over the limit.

Period 3: 37 – 61 hours

For the next 25 hours, the instance uses 2% CPU, which requires 30 credits. In the same period, it earns 75 credits, but the credit balance decreases. The balance decreases because the accrued launch credits are spent first, while newly earned credits are discarded because the credit balance is already at its limit of 72 earned credits.
### Credit Spend Rate
28.8 credits per 24 hours (1.2 credits per hour, 2% CPU utilization, 40% of credit earn rate)—30 credits over 25 hours

### Credit Earn Rate
72 credits per 24 hours

### Credit Discard Rate
72 credits per 24 hours (100% of credit earn rate)

### Credit Balance
72 credits (30 launch credits were spent; 72 earned credits remain unspent)

---

**Conclusion**

An instance spends launch credits first, before spending earned credits. Launch credits do not count towards the credit limit. After the launch credits are spent, the balance can never go higher than what can be earned in 24 hours. Furthermore, while an instance is running, it cannot get more launch credits.

**Period 4: 62 – 72 hours**

For the next 11 hours, the instance uses 2% CPU, which requires 13.2 credits. This is the same CPU utilization as in the previous period, but the balance does not decrease. It stays at 72 credits.

The balance does not decrease because the credit earn rate is higher than the credit spend rate. In the time that the instance spends 13.2 credits, it also earns 33 credits. However, the balance limit is 72 credits, so any earned credits that exceed the limit are discarded. The balance plateaus at 72 credits, which is different from the plateau of 102 credits during Period 2, because there are no accrued launch credits.
**Credit Spend Rate**

28.8 credits per 24 hours (1.2 credits per hour, 2% CPU utilization, 40% of credit earn rate)—13.2 credits over 11 hours

**Credit Earn Rate**

72 credits per 24 hours

**Credit Discard Rate**

43.2 credits per 24 hours (60% of credit earn rate)

**Credit Balance**

72 credits (0 launch credits, 72 earned credits)—balance is at its limit

**Conclusion**

After launch credits are spent, the credit balance limit is determined by the number of credits that an instance can earn in 24 hours. If the instance earns more credits than it spends, newly earned credits over the limit are discarded.

**Period 5: 73 – 75 hours**

For the next three hours, the instance bursts at 20% CPU utilization, which requires 36 credits. The instance earns nine credits in the same three hours, which results in a net balance decrease of 27 credits. At the end of three hours, the credit balance is 45 accrued earned credits.
Credit Spend Rate: 288 credits per 24 hours (12 credits per hour, 20% CPU utilization, 400% of credit earn rate)—36 credits over 3 hours.

Credit Earn Rate: 72 credits per 24 hours (9 credits over 3 hours).

Credit Discard Rate: 0 credits per 24 hours.

Credit Balance: 45 credits (previous balance (72) - spent credits (36) + earned credits (9))—balance decreases at a rate of 216 credits per 24 hours (spend rate 288/24 + earn rate 72/24 = balance decrease rate 216/24).

Conclusion

If an instance spends more credits than it earns, its credit balance decreases.

Period 6: 76 – 90 hours

For the next 15 hours, the instance uses 2% CPU, which requires 18 credits. This is the same CPU utilization as in Periods 3 and 4. However, the balance increases in this period, whereas it decreased in Period 3 and plateaued in Period 4.

In Period 3, the accrued launch credits were spent, and any earned credits that exceeded the credit limit were discarded, resulting in a decrease in the credit balance. In Period 4, the instance spent fewer credits than it earned. Any earned credits that exceeded the limit were discarded, so the balance plateaued at its maximum of 72 credits.
In this period, there are no accrued launch credits, and the number of accrued earned credits in the balance is below the limit. No earned credits are discarded. Furthermore, the instance earns more credits than it spends, resulting in an increase in the credit balance.

<table>
<thead>
<tr>
<th>Credit Spend Rate</th>
<th>28.8 credits per 24 hours (1.2 credits per hour, 2% CPU utilization, 40% of credit earn rate)—18 credits over 15 hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit Earn Rate</td>
<td>72 credits per 24 hours (45 credits over 15 hours)</td>
</tr>
<tr>
<td>Credit Discard Rate</td>
<td>0 credits per 24 hours</td>
</tr>
<tr>
<td>Credit Balance</td>
<td>72 credits (balance increases at a rate of 43.2 credits per 24 hours—change rate = spend rate 28.8/24 + earn rate 72/24)</td>
</tr>
</tbody>
</table>

**Conclusion**

If an instance spends fewer credits than it earns, its credit balance increases.

**Period 7: 91 – 96 hours**

For the next six hours, the instance remains idle—CPU utilization is 0%—and no credits are spent. This is the same CPU utilization as in Period 2, but the balance does not plateau at 102 credits—it plateaus at 72 credits, which is the credit balance limit for the instance.

In Period 2, the credit balance included 30 accrued launch credits. The launch credits were spent in Period 3. A running instance cannot get more launch credits. After its credit balance limit is reached, any earned credits that exceed the limit are discarded.
Credit Spend Rate 0 credits per 24 hours (0% CPU utilization)
Credit Earn Rate 72 credits per 24 hours
Credit Discard Rate 72 credits per 24 hours (100% of credit earn rate)
Credit Balance 72 credits (0 launch credits, 72 earned credits)

Conclusion
An instance constantly earns credits, but cannot accrue more earned credits if the credit balance limit has been reached. After the limit is reached, newly earned credits are discarded. The credit balance limit is determined by the number of credits that an instance can earn in 24 hours. For more information about credit balance limits, see the credit table (p. 133).

Working with Burstable Performance Instances
The steps for launching, monitoring, and modifying these instances are similar. The key difference is the default credit specification when they launch:

• T3 and T3a instances launch as unlimited by default.
• T2 instances launch as standard by default.

Contents
• Launching a Burstable Performance Instance as Unlimited or Standard (p. 154)
• Using an Auto Scaling Group to Launch a Burstable Performance Instance as Unlimited (p. 154)
• Viewing the Credit Specification of a Burstable Performance Instance (p. 156)
• Modifying the Credit Specification of a Burstable Performance Instance (p. 156)

Launching a Burstable Performance Instance as Unlimited or Standard

T3 and T3a instances launch as unlimited by default. T2 instances launch as standard by default.

You can launch your instances as unlimited or standard using the Amazon EC2 console, an AWS SDK, a command line tool, or with an Auto Scaling group. For more information, see Using an Auto Scaling Group to Launch a Burstable Performance Instance as Unlimited (p. 154).

To launch a burstable performance instance as Unlimited or Standard (console)

1. Follow the Launching an Instance Using the Launch Instance Wizard (p. 334) procedure.
2. On the Choose an Instance Type page, select an instance type, and choose Next: Configure Instance Details.
3. Choose a credit specification. The default for T3 and T3a is unlimited, and for T2 it is standard.
   a. To launch a T3 or T3a instance as standard, on the Configure Instance Details page, for T2/T3 Unlimited, clear Enable.
   b. To launch a T2 instance as unlimited, on the Configure Instance Details page, for T2/T3 Unlimited, select Enable.
4. Continue as prompted by the wizard. When you've finished reviewing your options on the Review Instance Launch page, choose Launch. For more information, see Launching an Instance Using the Launch Instance Wizard (p. 334).

To launch a burstable performance instance as Unlimited or Standard (AWS CLI)

Use the run-instances command to launch your instances. Specify the credit specification using the --credit-specification CpuCredits= parameter. Valid credit specifications are unlimited and standard.

• For T3 and T3a, if you do not include the --credit-specification parameter, the instance launches as unlimited by default.
• For T2, if you do not include the --credit-specification parameter, the instance launches as standard by default.

aws ec2 run-instances --image-id ami-abc12345 --count 1 --instance-type t3.micro --key-name MyKeyPair --credit-specification "CpuCredits=unlimited"

Using an Auto Scaling Group to Launch a Burstable Performance Instance as Unlimited

When burstable performance instances are launched or started, they require CPU credits for a good bootstrapping experience. If you use an Auto Scaling group to launch your instances, we recommend that you configure your instances as unlimited. If you do, the instances use surplus credits when they are automatically launched or restarted by the Auto Scaling group. Using surplus credits prevents performance restrictions.

Creating a Launch Template

You must use a launch template for launching instances as unlimited in an Auto Scaling group. A launch configuration does not support launching instances as unlimited.

To create a launch template that launches instances as Unlimited (console)

1. Follow the Creating a Launch Template for an Auto Scaling Group procedure.
2. In **Launch template contents**, for **Instance type**, choose a T3, T3a, or T2 instance size.

3. To launch instances as unlimited in an Auto Scaling group, in **Advanced details**, for **T2/T3 Unlimited**, choose **Enable**.

4. When you've finished defining the launch template parameters, choose **Create launch template**. For more information, see **Creating a Launch Template for an Auto Scaling Group** in the *Amazon EC2 Auto Scaling User Guide*.

**To create a launch template that launches instances as Unlimited (AWS CLI)**

Use the `create-launch-template` command and specify `unlimited` as the credit specification.

- For T3 and T3a, if you do not include the `CreditSpecification={CpuCredits=unlimited}` value, the instance launches as unlimited by default.
- For T2, if you do not include the `CreditSpecification={CpuCredits=unlimited}` value, the instance launches as standard by default.

```bash
aws ec2 create-launch-template --launch-template-name MyLaunchTemplate --version-description FirstVersion --launch-template-data ImageId=ami-8c1be5f6,InstanceType=t3.medium,CreditSpecification={CpuCredits=unlimited}
```

**Associating an Auto Scaling Group with a Launch Template**

To associate the launch template with an Auto Scaling group, create the Auto Scaling group using the launch template, or add the launch template to an existing Auto Scaling group.

**To create an Auto Scaling group using a launch template (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the navigation bar at the top of the screen, select the same Region that you used when you created the launch template.
3. In the navigation pane, choose **Auto Scaling Groups**, **Create Auto Scaling group**.
4. Choose **Launch Template**, select your launch template, and then choose **Next Step**.
5. Complete the fields for the Auto Scaling group. When you've finished reviewing your configuration settings on the **Review page**, choose **Create Auto Scaling group**. For more information, see **Creating an Auto Scaling Group Using a Launch Template** in the *Amazon EC2 Auto Scaling User Guide*.

**To create an Auto Scaling group using a launch template (AWS CLI)**

Use the `create-auto-scaling-group` AWS CLI command and specify the `--launch-template` parameter.

**To add a launch template to an existing Auto Scaling group (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the navigation bar at the top of the screen, select the same Region that you used when you created the launch template.
3. In the navigation pane, choose **Auto Scaling Groups**.
4. From the Auto Scaling group list, select an Auto Scaling group, and choose **Actions, Edit**.
5. On the **Details** tab, for **Launch Template**, choose a launch template, and then choose **Save**.

**To add a launch template to an existing Auto Scaling group (AWS CLI)**

Use the `update-auto-scaling-group` AWS CLI command and specify the `--launch-template` parameter.
Viewing the Credit Specification of a Burstable Performance Instance

You can view the credit specification (unlimited or standard) of a running or stopped instance.

To view the credit specification of a burstable instance (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the left navigation pane, choose Instances and select the instance.
3. Choose Description and view the T2/T3 Unlimited field.
   - If the value is Enabled, then your instance is configured as unlimited.
   - If the value is Disabled, then your instance is configured as standard.

To describe the credit specification of a burstable performance instance (AWS CLI)

Use the describe-instance-credit-specifications command. If you do not specify one or more instance IDs, all instances with the credit specification of unlimited are returned, as well as instances that were previously configured with the unlimited credit specification. For example, if you resize a T3 instance to an M4 instance, while it is configured as unlimited, Amazon EC2 returns the M4 instance.

Example

```bash
aws ec2 describe-instance-credit-specifications --instance-id i-1234567890abcdef0
```

The following is example output:

```
{
   "InstanceCreditSpecifications": [
      {
         "InstanceId": "i-1234567890abcdef0",
         "CpuCredits": "unlimited"
      }
   ]
}
```

Modifying the Credit Specification of a Burstable Performance Instance

You can switch the credit specification of a running or stopped instance at any time between unlimited and standard.

To modify the credit specification of a burstable performance instance (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the left navigation pane, choose Instances and select the instance. To modify the credit specification for several instances at one time, select all applicable instances.
3. Choose Actions, Instance Settings, Change T2/T3 Unlimited.
   - Note: The Change T2/T3 Unlimited option is enabled only if you select a T3, T3a, or T2 instance.
4. To change the credit specification to unlimited, choose Enable. To change the credit specification to standard, choose Disable. The current credit specification for the instance appears in parentheses after the instance ID.

To modify the credit specification of a burstable performance instance (AWS CLI)
Use the `modify-instance-credit-specification` command. Specify the instance and its credit specification using the `--instance-credit-specification` parameter. Valid credit specifications are `unlimited` and `standard`.

**Example**

```bash
aws ec2 modify-instance-credit-specification --region us-east-1 --instance-credit-specification "InstanceId=i-1234567890abcdef0,CpuCredits=unlimited"
```

The following is example output:

```json
{
  "SuccessfulInstanceCreditSpecifications": [
    {
      "InstanceId": "i-1234567890abcdef0"
    }
  ],
  "UnsuccessfulInstanceCreditSpecifications": []
}
```

**Monitoring Your CPU Credits**

You can see the credit balance for each instance in the Amazon EC2 per-instance metrics of the CloudWatch console.

**Topics**

- Additional CloudWatch Metrics for Burstable Performance Instances (p. 157)
- Calculating CPU Credit Usage (p. 159)

**Additional CloudWatch Metrics for Burstable Performance Instances**

T3, T3a, and T2 instances have these additional CloudWatch metrics, which are updated every five minutes:

- **CPUCreditUsage** – The number of CPU credits spent during the measurement period.
- **CPUCreditBalance** – The number of CPU credits that an instance has accrued. This balance is depleted when the CPU bursts and CPU credits are spent more quickly than they are earned.
- **CPUSurplusCreditBalance** – The number of surplus CPU credits spent to sustain CPU performance when the `CPUCreditBalance` value is zero.
- **CPUSurplusCreditsCharged** – The number of surplus CPU credits exceeding the maximum number of CPU credits (p. 133) that can be earned in a 24-hour period, and thus attracting an additional charge.

The last two metrics apply only to instances configured as `unlimited`.

The following table describes the CloudWatch metrics for burstable performance instances. For more information, see List the Available CloudWatch Metrics for Your Instances (p. 548).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUCreditUsage</td>
<td>The number of CPU credits spent by the instance for CPU utilization. One CPU credit equals one vCPU running at 100% utilization for one minute or an equivalent combination of vCPUs, utilization, and time (for example, one vCPU running at 50%</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>CPUCreditBalance</td>
<td>The number of earned CPU credits that an instance has accrued since it was launched or started. For T2 Standard, the CPUCreditBalance also includes the number of launch credits that have been accrued.</td>
</tr>
<tr>
<td></td>
<td>Credits are accrued in the credit balance after they are earned, and removed from the credit balance when they are spent. The credit balance has a maximum limit, determined by the instance size. After the limit is reached, any new credits that are earned are discarded. For T2 Standard, launch credits do not count towards the limit.</td>
</tr>
<tr>
<td></td>
<td>The credits in the CPUCreditBalance are available for the instance to spend to burst beyond its baseline CPU utilization.</td>
</tr>
<tr>
<td></td>
<td>When an instance is running, credits in the CPUCreditBalance do not expire. When a T3 or T3a instance stops, the CPUCreditBalance value persists for seven days. Thereafter, all accrued credits are lost. When a T2 instance stops, the CPUCreditBalance value does not persist, and all accrued credits are lost.</td>
</tr>
<tr>
<td></td>
<td>CPU credit metrics are available at a five-minute frequency only.</td>
</tr>
<tr>
<td></td>
<td>Units: Credits (vCPU-minutes)</td>
</tr>
<tr>
<td>CPUSurplusCreditBalance</td>
<td>The number of surplus credits that have been spent by an unlimited instance when its CPUCreditBalance value is zero.</td>
</tr>
<tr>
<td></td>
<td>The CPUSurplusCreditBalance value is paid down by earned CPU credits. If the number of surplus credits exceeds the maximum number of credits that the instance can earn in a 24-hour period, the spent surplus credits above the maximum incur an additional charge.</td>
</tr>
<tr>
<td></td>
<td>Units: Credits (vCPU-minutes)</td>
</tr>
</tbody>
</table>
### Metric Description

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUSurplusCreditsCharged</td>
<td>The number of spent surplus credits that are not paid down by earned CPU credits, and which thus incur an additional charge.</td>
</tr>
</tbody>
</table>

Spent surplus credits are charged when any of the following occurs:

- The spent surplus credits exceed the maximum number of credits that the instance can earn in a 24-hour period. Spent surplus credits above the maximum are charged at the end of the hour.
- The instance is stopped or terminated.
- The instance is switched from unlimited to standard.

Units: Credits (vCPU-minutes)

### Calculating CPU Credit Usage

The CPU credit usage of instances is calculated using the instance CloudWatch metrics described in the preceding table.

Amazon EC2 sends the metrics to CloudWatch every five minutes. A reference to the prior value of a metric at any point in time implies the previous value of the metric, sent five minutes ago.

### Calculating CPU Credit Usage for Standard Instances

- The CPU credit balance increases if CPU utilization is below the baseline, when the credits spent are less than the credits earned in the prior five-minute interval.
- The CPU credit balance decreases if CPU utilization is above the baseline, when the credits spent are more than the credits earned in the prior five-minute interval.

Mathematically, this is captured by the following equation:

\[
\text{CPUCreditBalance} = \text{prior CPUCreditBalance} + [\text{Credits earned per hour} \times (5/60) - \text{CPUCreditUsage}]
\]

The size of the instance determines the number of credits that the instance can earn per hour and the number of earned credits that it can accrue in the credit balance. For information about the number of credits earned per hour, and the credit balance limit for each instance size, see the credit table (p. 133).

### Example

This example uses a t3.nano instance. To calculate the CPUCreditBalance value of the instance, use the preceding equation as follows:

- CPUCreditBalance – The current credit balance to calculate.
- prior CPUCreditBalance – The credit balance five minutes ago. In this example, the instance had accrued two credits.
- Credits earned per hour – A t3.nano instance earns six credits per hour.
- 5/60 – Represents the five-minute interval between CloudWatch metric publication. Multiply the credits earned per hour by 5/60 (five minutes) to get the number of credits that the instance earned in the past five minutes. A t3.nano instance earns 0.5 credits every five minutes.
• **CPUCreditUsage** – How many credits the instance spent in the past five minutes. In this example, the instance spent one credit in the past five minutes.

Using these values, you can calculate the **CPUCreditBalance** value:

**Example**

![Equation](equation.png)

**Calculating CPU Credit Usage for Unlimited Instances**

When a T3, T3a, or T2 instance needs to burst above the baseline, it always spends accrued credits before spending surplus credits. When it depletes its accrued CPU credit balance, it can spend surplus credits to burst for as long as it needs. When CPU utilization falls below the baseline, surplus credits are always paid down before the instance accrues earned credits.

We use the term **Adjusted balance** in the following equations to reflect the activity that occurs in this five-minute interval. We use this value to arrive at the values for the **CPUCreditBalance** and **CPUSurplusCreditBalance** CloudWatch metrics.

**Example**

![Equation](equation.png)

A value of 0 for **Adjusted balance** indicates that the instance spent all its earned credits for bursting, and no surplus credits were spent. As a result, both **CPUCreditBalance** and **CPUSurplusCreditBalance** are set to 0.

A positive **Adjusted balance** value indicates that the instance accrued earned credits, and previous surplus credits, if any, were paid down. As a result, the **Adjusted balance** value is assigned to **CPUCreditBalance**, and the **CPUSurplusCreditBalance** is set to 0. The instance size determines the maximum number of credits (p. 133) that it can accrue.

**Example**

![Equation](equation.png)

A negative **Adjusted balance** value indicates that the instance spent all its earned credits that it accrued and, in addition, also spent surplus credits for bursting. As a result, the **Adjusted balance** value is assigned to **CPUSurplusCreditBalance** and **CPUCreditBalance** is set to 0. Again, the instance size determines the maximum number of credits (p. 133) that it can accrue.

**Example**

![Equation](equation.png)

If the surplus credits spent exceed the maximum credits that the instance can accrue, the surplus credit balance is set to the maximum, as shown in the preceding equation. The remaining surplus credits are charged as represented by the **CPUSurplusCreditsCharged** metric.

**Example**

![Equation](equation.png)
Finally, when the instance terminates, any surplus credits tracked by the CPUSurplusCreditBalance are charged. If the instance is switched from unlimited to standard, any remaining CPUSurplusCreditBalance is also charged.

## Compute Optimized Instances

Compute optimized instances are ideal for compute-bound applications that benefit from high-performance processors. They are well suited for the following applications:

- Batch processing workloads
- Media transcoding
- High-performance web servers
- High-performance computing (HPC)
- Scientific modeling
- Dedicated gaming servers and ad serving engines
- Machine learning inference and other compute-intensive applications

For more information, see Amazon EC2 C5 Instances.

### Contents

- Hardware Specifications (p. 161)
- Instance Performance (p. 162)
- Network Performance (p. 162)
- SSD I/O Performance (p. 163)
- Instance Features (p. 164)
- Release Notes (p. 164)

## Hardware Specifications

The following is a summary of the hardware specifications for compute optimized instances.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Memory (GiB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c4.large</td>
<td>2</td>
<td>3.75</td>
</tr>
<tr>
<td>c4.xlarge</td>
<td>4</td>
<td>7.5</td>
</tr>
<tr>
<td>c4.2xlarge</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>c4.4xlarge</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>c4.8xlarge</td>
<td>36</td>
<td>60</td>
</tr>
<tr>
<td>c5.large</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>c5.xlarge</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>c5.2xlarge</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>c5.4xlarge</td>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>c5.9xlarge</td>
<td>36</td>
<td>72</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Memory (GiB)</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>c5.12xlarge</td>
<td>48</td>
<td>96</td>
</tr>
<tr>
<td>c5.18xlarge</td>
<td>72</td>
<td>144</td>
</tr>
<tr>
<td>c5.24xlarge</td>
<td>96</td>
<td>192</td>
</tr>
<tr>
<td>c5.metal</td>
<td>96</td>
<td>192</td>
</tr>
<tr>
<td>c5d.large</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>c5d.xlarge</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>c5d.2xlarge</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>c5d.4xlarge</td>
<td>16</td>
<td>32</td>
</tr>
<tr>
<td>c5d.9xlarge</td>
<td>36</td>
<td>72</td>
</tr>
<tr>
<td>c5d.18xlarge</td>
<td>72</td>
<td>144</td>
</tr>
<tr>
<td>c5n.large</td>
<td>2</td>
<td>5.25</td>
</tr>
<tr>
<td>c5n.xlarge</td>
<td>4</td>
<td>10.5</td>
</tr>
<tr>
<td>c5n.2xlarge</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>c5n.4xlarge</td>
<td>16</td>
<td>42</td>
</tr>
<tr>
<td>c5n.9xlarge</td>
<td>36</td>
<td>96</td>
</tr>
<tr>
<td>c5n.18xlarge</td>
<td>72</td>
<td>192</td>
</tr>
<tr>
<td>c5n.metal</td>
<td>72</td>
<td>192</td>
</tr>
</tbody>
</table>

For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.

For more information about specifying CPU options, see Optimizing CPU Options (p. 447).

**Instance Performance**

EBS-optimized instances enable you to get consistently high performance for your EBS volumes by eliminating contention between Amazon EBS I/O and other network traffic from your instance. Some compute optimized instances are EBS-optimized by default at no additional cost. For more information, see Amazon EBS-Optimized Instances (p. 876).

**Network Performance**

You can enable enhanced networking capabilities on supported instance types. Enhanced networking provides significantly higher packet-per-second (PPS) performance, lower network jitter, and lower latencies. For more information, see Enhanced Networking on Windows (p. 700).

Instance types that use the Elastic Network Adapter (ENA) for enhanced networking deliver high packet per second performance with consistently low latencies. Most applications do not consistently need a high level of network performance, but can benefit from having access to increased bandwidth when they send or receive data. Instance sizes that use the ENA and are documented with network performance of "Up to 10 Gbps" or "Up to 25 Gbps" use a network I/O credit mechanism to allocate network bandwidth to instances based on average bandwidth utilization. These instances accrue credits
when their network bandwidth is below their baseline limits, and can use these credits when they perform network data transfers.

The following is a summary of network performance for compute optimized instances that support enhanced networking.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Network performance</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>c5.4xlarge and smaller</td>
<td>Up to 10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5d.4xlarge and smaller</td>
<td>Up to 10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5.9xlarge</td>
<td>10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5d.9xlarge</td>
<td>10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5.12xlarge</td>
<td>12 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5d.12xlarge</td>
<td>12 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5n.4xlarge and smaller</td>
<td>Up to 25 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5.18xlarge</td>
<td>25 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5d.18xlarge</td>
<td>25 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5n.9xlarge</td>
<td>50 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c5n.18xlarge</td>
<td>100 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>c4.large</td>
<td>Moderate</td>
<td>Intel 82599 VF (p. 707)</td>
</tr>
<tr>
<td>c4.xlarge</td>
<td>High</td>
<td>Intel 82599 VF (p. 707)</td>
</tr>
<tr>
<td>c4.2xlarge</td>
<td>High</td>
<td>Intel 82599 VF (p. 707)</td>
</tr>
<tr>
<td>c4.4xlarge</td>
<td>High</td>
<td>Intel 82599 VF (p. 707)</td>
</tr>
<tr>
<td>c4.8xlarge</td>
<td>10 Gbps</td>
<td>Intel 82599 VF (p. 707)</td>
</tr>
</tbody>
</table>

SSD I/O Performance

If you use all the SSD-based instance store volumes available to your instance, you get the IOPS (4,096 byte block size) performance listed in the following table (at queue depth saturation). Otherwise, you get lower IOPS performance.

<table>
<thead>
<tr>
<th>Instance Size</th>
<th>100% Random Read IOPS</th>
<th>Write IOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>c5d.large *</td>
<td>20,000</td>
<td>9,000</td>
</tr>
<tr>
<td>c5d.xlarge *</td>
<td>40,000</td>
<td>18,000</td>
</tr>
<tr>
<td>c5d.2xlarge *</td>
<td>80,000</td>
<td>37,000</td>
</tr>
<tr>
<td>c5d.4xlarge *</td>
<td>175,000</td>
<td>75,000</td>
</tr>
<tr>
<td>c5d.9xlarge</td>
<td>350,000</td>
<td>170,000</td>
</tr>
<tr>
<td>c5d.18xlarge</td>
<td>700,000</td>
<td>340,000</td>
</tr>
</tbody>
</table>

* For these instances, you can get up to the specified performance.

As you fill the SSD-based instance store volumes for your instance, the number of write IOPS that you can achieve decreases. This is due to the extra work the SSD controller must do to find available space, rewrite existing data, and erase unused space so that it can be rewritten. This process of garbage collection results in internal write amplification to the SSD, expressed as the ratio of SSD write
operations to user write operations. This decrease in performance is even larger if the write operations are not in multiples of 4,096 bytes or not aligned to a 4,096-byte boundary. If you write a smaller amount of bytes or bytes that are not aligned, the SSD controller must read the surrounding data and store the result in a new location. This pattern results in significantly increased write amplification, increased latency, and dramatically reduced I/O performance.

SSD controllers can use several strategies to reduce the impact of write amplification. One such strategy is to reserve space in the SSD instance storage so that the controller can more efficiently manage the space available for write operations. This is called **over-provisioning**. The SSD-based instance store volumes provided to an instance don’t have any space reserved for over-provisioning. To reduce write amplification, we recommend that you leave 10% of the volume unpartitioned so that the SSD controller can use it for over-provisioning. This decreases the storage that you can use, but increases performance even if the disk is close to full capacity.

For instance store volumes that support TRIM, you can use the TRIM command to notify the SSD controller whenever you no longer need data that you’ve written. This provides the controller with more free space, which can reduce write amplification and increase performance. For more information, see Instance Store Volume TRIM Support (p. 925).

**Instance Features**

The following is a summary of features for compute optimized instances:

<table>
<thead>
<tr>
<th>EBS only</th>
<th>NVMe EBS</th>
<th>Instance store</th>
<th>Placement group</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>C5</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>C5d</td>
<td>No</td>
<td>Yes</td>
<td>NVMe *</td>
</tr>
<tr>
<td>C5n</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

* The root device volume must be an Amazon EBS volume.

For more information, see the following:

- Amazon EBS and NVMe on Windows Instances (p. 875)
- Amazon EC2 Instance Store (p. 917)
- Placement Groups (p. 710)

**Release Notes**

- C4, C5, C5d, and C5n instances require 64-bit EBS-backed HVM AMIs. They have high-memory and require a 64-bit operating system to take advantage of that capacity. HVM AMIs provide superior performance in comparison to paravirtual (PV) AMIs on high-memory instance types. In addition, you must use an HVM AMI to take advantage of enhanced networking.
- C5, C5d, and C5n instances have the following requirements:
  - NVMe drivers must be installed. EBS volumes are exposed as NVMe block devices (p. 875).
  - Elastic Network Adapter (ENA (p. 701)) drivers must be installed.

The following AMIs meet these requirements:

- Amazon Linux 2
- Amazon Linux AMI 2018.03
Memory Optimized Instances

Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory.

**R4, R5, R5a, R5ad, R5d Instances**

These instances are well suited for the following applications:

- High-performance, relational (MySQL) and NoSQL (MongoDB, Cassandra) databases.
- Distributed web scale cache stores that provide in-memory caching of key-value type data (Memcached and Redis).
- In-memory databases using optimized data storage formats and analytics for business intelligence (for example, SAP HANA).
- Applications performing real-time processing of big unstructured data (financial services, Hadoop/Spark clusters).
- High-performance computing (HPC) and Electronic Design Automation (EDA) applications.

**r5.metal and r5d.metal instances** provide your applications with direct access to physical resources of the host server, such as processors and memory. These instances are well suited for the following:

- Workloads that require access to low-level hardware features (for example, Intel VT) that are not available or fully supported in virtualized environments
- Applications that require a non-virtualized environment for licensing or support

For more information, see Amazon EC2 R5 Instances.

**High memory instances**

High memory instances (u-6tb1.metal, u-9tb1.metal, and u-12tb1.metal) offer 6 TiB, 9 TiB, and 12 TiB of memory per instance. These instances are designed to run large in-memory databases,
including production installations of SAP HANA. They offer bare metal performance with direct access to host hardware.

**X1 Instances**

These instances are well suited for the following applications:

- In-memory databases such as SAP HANA, including SAP-certified support for Business Suite S/4HANA, Business Suite on HANA (SoH), Business Warehouse on HANA (BW), and Data Mart Solutions on HANA. For more information, see [SAP HANA on the AWS Cloud](https://aws.amazon.com/hana/).
- Big-data processing engines such as Apache Spark or Presto.
- High-performance computing (HPC) applications.

For more information, see [Amazon EC2 X1 Instances](https://aws.amazon.com/ec2/x1/).

**X1e Instances**

These instances are well suited for the following applications:

- High-performance databases.
- In-memory databases such as SAP HANA. For more information, see [SAP HANA on the AWS Cloud](https://aws.amazon.com/hana/).
- Memory-intensive enterprise applications.

For more information, see [Amazon EC2 X1e Instances](https://aws.amazon.com/ec2/x1e/).

**z1d Instances**

These instances deliver both high compute and high memory and are well-suited for the following applications:

- Electronic Design Automation (EDA)
- Relational database workloads

z1d.metal instances provide your applications with direct access to physical resources of the host server, such as processors and memory. These instances are well suited for the following:

- Workloads that require access to low-level hardware features (for example, Intel VT) that are not available or fully supported in virtualized environments
- Applications that require a non-virtualized environment for licensing or support

For more information, see [Amazon EC2 z1d Instances](https://aws.amazon.com/ec2/z1d/).

**Contents**

- Hardware Specifications (p. 167)
- Memory Performance (p. 169)
- Instance Performance (p. 169)
- Network Performance (p. 169)
- SSD I/O Performance (p. 170)
- Instance Features (p. 171)
- High Availability and Reliability (X1) (p. 172)
- Support for vCPUs (p. 172)
Hardware Specifications

The following is a summary of the hardware specifications for memory optimized instances.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Memory (GiB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>r4.large</td>
<td>2</td>
<td>15.25</td>
</tr>
<tr>
<td>r4.xlarge</td>
<td>4</td>
<td>30.5</td>
</tr>
<tr>
<td>r4.2xlarge</td>
<td>8</td>
<td>61</td>
</tr>
<tr>
<td>r4.4xlarge</td>
<td>16</td>
<td>122</td>
</tr>
<tr>
<td>r4.8xlarge</td>
<td>32</td>
<td>244</td>
</tr>
<tr>
<td>r4.16xlarge</td>
<td>64</td>
<td>488</td>
</tr>
<tr>
<td>r5.large</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>r5.xlarge</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>r5.2xlarge</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>r5.4xlarge</td>
<td>16</td>
<td>128</td>
</tr>
<tr>
<td>r5.8xlarge</td>
<td>32</td>
<td>256</td>
</tr>
<tr>
<td>r5.12xlarge</td>
<td>48</td>
<td>384</td>
</tr>
<tr>
<td>r5.16xlarge</td>
<td>64</td>
<td>512</td>
</tr>
<tr>
<td>r5.24xlarge</td>
<td>96</td>
<td>768</td>
</tr>
<tr>
<td>r5.metal</td>
<td>96</td>
<td>768</td>
</tr>
<tr>
<td>r5a.large</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>r5a.xlarge</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>r5a.2xlarge</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>r5a.4xlarge</td>
<td>16</td>
<td>128</td>
</tr>
<tr>
<td>r5a.8xlarge</td>
<td>32</td>
<td>256</td>
</tr>
<tr>
<td>r5a.12xlarge</td>
<td>48</td>
<td>384</td>
</tr>
<tr>
<td>r5a.16xlarge</td>
<td>64</td>
<td>512</td>
</tr>
<tr>
<td>r5a.24xlarge</td>
<td>96</td>
<td>768</td>
</tr>
<tr>
<td>r5ad.large</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>r5ad.xlarge</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>r5ad.2xlarge</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>r5ad.4xlarge</td>
<td>16</td>
<td>128</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Memory (GiB)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>r5ad.12xlarge</td>
<td>48</td>
<td>384</td>
</tr>
<tr>
<td>r5ad.24xlarge</td>
<td>96</td>
<td>768</td>
</tr>
<tr>
<td>r5d.large</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>r5d.xlarge</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>r5d.2xlarge</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>r5d.4xlarge</td>
<td>16</td>
<td>128</td>
</tr>
<tr>
<td>r5d.8xlarge</td>
<td>32</td>
<td>256</td>
</tr>
<tr>
<td>r5d.12xlarge</td>
<td>48</td>
<td>384</td>
</tr>
<tr>
<td>r5d.16xlarge</td>
<td>64</td>
<td>512</td>
</tr>
<tr>
<td>r5d.24xlarge</td>
<td>96</td>
<td>768</td>
</tr>
<tr>
<td>r5d.metal</td>
<td>96</td>
<td>768</td>
</tr>
<tr>
<td>u-6tb1.metal</td>
<td>448 *</td>
<td>6,144</td>
</tr>
<tr>
<td>u-9tb1.metal</td>
<td>448 *</td>
<td>9,216</td>
</tr>
<tr>
<td>u-12tb1.metal</td>
<td>448 *</td>
<td>12,288</td>
</tr>
<tr>
<td>x1.16xlarge</td>
<td>64</td>
<td>976</td>
</tr>
<tr>
<td>x1.32xlarge</td>
<td>128</td>
<td>1,952</td>
</tr>
<tr>
<td>x1e.xlarge</td>
<td>4</td>
<td>122</td>
</tr>
<tr>
<td>x1e.2xlarge</td>
<td>8</td>
<td>244</td>
</tr>
<tr>
<td>x1e.4xlarge</td>
<td>16</td>
<td>488</td>
</tr>
<tr>
<td>x1e.8xlarge</td>
<td>32</td>
<td>976</td>
</tr>
<tr>
<td>x1e.16xlarge</td>
<td>64</td>
<td>1,952</td>
</tr>
<tr>
<td>x1e.32xlarge</td>
<td>128</td>
<td>3,904</td>
</tr>
<tr>
<td>z1d.large</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>z1d.xlarge</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>z1d.2xlarge</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>z1d.3xlarge</td>
<td>12</td>
<td>96</td>
</tr>
<tr>
<td>z1d.6xlarge</td>
<td>24</td>
<td>192</td>
</tr>
<tr>
<td>z1d.12xlarge</td>
<td>48</td>
<td>384</td>
</tr>
<tr>
<td>z1d.metal</td>
<td>48</td>
<td>384</td>
</tr>
</tbody>
</table>

* Each logical processor is a hyperthread on 224 cores.
For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.

For more information about specifying CPU options, see Optimizing CPU Options (p. 447).

**Memory Performance**

X1 instances include Intel Scalable Memory Buffers, providing 300 GiB/s of sustainable memory-read bandwidth and 140 GiB/s of sustainable memory-write bandwidth.

For more information about how much RAM can be enabled for memory optimized instances, see Hardware Specifications (p. 167).

Memory optimized instances have high memory and require 64-bit HVM AMIs to take advantage of that capacity. HVM AMIs provide superior performance in comparison to paravirtual (PV) AMIs on memory optimized instances.

**Instance Performance**

R4 instances feature up to 64 vCPUs and are powered by two AWS-customized Intel XEON processors based on E5-2686v4 that feature high-memory bandwidth and larger L3 caches to boost the performance of in-memory applications.

X1e and X1 instances feature up to 128 vCPUs and are powered by four Intel Xeon E7-8880 v3 processors that feature high-memory bandwidth and larger L3 caches to boost the performance of in-memory applications.

High memory instances (u-6tb1.metal, u-9tb1.metal, and u-12tb1.metal) are the first instances to be powered by an eight-socket platform with the latest generation Intel Xeon Platinum 8176M (Skylake) processors that are optimized for mission-critical enterprise workloads.

Memory optimized instances enable increased cryptographic performance through the latest Intel AES-NI feature, support Intel Transactional Synchronization Extensions (TSX) to boost the performance of in-memory transactional data processing, and support Advanced Vector Extensions 2 (Intel AVX2) processor instructions to expand most integer commands to 256 bits.

**Network Performance**

You can enable enhanced networking capabilities on supported instance types. Enhanced networking provides significantly higher packet-per-second (PPS) performance, lower network jitter, and lower latencies. For more information, see Enhanced Networking on Windows (p. 700).

Instance types that use the Elastic Network Adapter (ENA) for enhanced networking deliver high packet per second performance with consistently low latencies. Most applications do not consistently need a high level of network performance, but can benefit from having access to increased bandwidth when they send or receive data. Instance sizes that use the ENA and are documented with network performance of "Up to 10 Gbps" or "Up to 25 Gbps" use a network I/O credit mechanism to allocate network bandwidth to instances based on average bandwidth utilization. These instances accrue credits when their network bandwidth is below their baseline limits, and can use these credits when they perform network data transfers.

The following is a summary of network performance for memory optimized instances that support enhanced networking.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Network performance</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>r4.4xlarge and smaller</td>
<td>Up to 10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>r5.4xlarge and smaller</td>
<td>Up to 10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>r5a.8xlarge and smaller</td>
<td>Up to 10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
</tbody>
</table>
### Instance Type

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Network performance</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>r5ad.4xlarge and smaller</td>
<td>r5d.4xlarge and smaller</td>
<td>x1e.8xlarge and smaller</td>
</tr>
<tr>
<td>r4.8xlarge</td>
<td>r5.8xlarge</td>
<td>r5.12xlarge</td>
</tr>
<tr>
<td>r5.16xlarge</td>
<td>r5a.24xlarge</td>
<td>r5ad.24xlarge</td>
</tr>
<tr>
<td>r4.16xlarge</td>
<td>r5.24xlarge</td>
<td>r5.metal</td>
</tr>
</tbody>
</table>

### SSD I/O Performance

If you use all the SSD-based instance store volumes available to your instance, you get the IOPS (4,096 byte block size) performance listed in the following table (at queue depth saturation). Otherwise, you get lower IOPS performance.

<table>
<thead>
<tr>
<th>Instance Size</th>
<th>100% Random Read IOPS</th>
<th>Write IOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>r5ad.large *</td>
<td>30,000</td>
<td>15,000</td>
</tr>
<tr>
<td>r5ad.xlarge *</td>
<td>59,000</td>
<td>29,000</td>
</tr>
<tr>
<td>r5ad.2xlarge *</td>
<td>117,000</td>
<td>57,000</td>
</tr>
<tr>
<td>r5ad.4xlarge *</td>
<td>234,000</td>
<td>114,000</td>
</tr>
<tr>
<td>r5ad.12xlarge</td>
<td>700,000</td>
<td>340,000</td>
</tr>
<tr>
<td>r5ad.24xlarge</td>
<td>1,400,000</td>
<td>680,000</td>
</tr>
<tr>
<td>r5d.large *</td>
<td>30,000</td>
<td>15,000</td>
</tr>
<tr>
<td>r5d.xlarge *</td>
<td>59,000</td>
<td>29,000</td>
</tr>
<tr>
<td>r5d.2xlarge *</td>
<td>117,000</td>
<td>57,000</td>
</tr>
<tr>
<td>r5d.4xlarge *</td>
<td>234,000</td>
<td>114,000</td>
</tr>
<tr>
<td>r5d.8xlarge</td>
<td>466,666</td>
<td>233,333</td>
</tr>
<tr>
<td>r5d.12xlarge</td>
<td>700,000</td>
<td>340,000</td>
</tr>
<tr>
<td>r5d.16xlarge</td>
<td>933,333</td>
<td>466,666</td>
</tr>
<tr>
<td>r5d.24xlarge</td>
<td>1,400,000</td>
<td>680,000</td>
</tr>
<tr>
<td>r5d.metal</td>
<td>1,400,000</td>
<td>680,000</td>
</tr>
</tbody>
</table>
### Instance Size

<table>
<thead>
<tr>
<th>Instance Size</th>
<th>100% Random Read IOPS</th>
<th>Write IOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>z1d.large *</td>
<td>30,000</td>
<td>15,000</td>
</tr>
<tr>
<td>z1d.xlarge *</td>
<td>59,000</td>
<td>29,000</td>
</tr>
<tr>
<td>z1d.2xlarge *</td>
<td>117,000</td>
<td>57,000</td>
</tr>
<tr>
<td>z1d.3xlarge *</td>
<td>175,000</td>
<td>75,000</td>
</tr>
<tr>
<td>z1d.6xlarge</td>
<td>350,000</td>
<td>170,000</td>
</tr>
<tr>
<td>z1d.12xlarge</td>
<td>700,000</td>
<td>340,000</td>
</tr>
<tr>
<td>z1d.metal</td>
<td>700,000</td>
<td>340,000</td>
</tr>
</tbody>
</table>

* For these instances, you can get up to the specified performance.

As you fill the SSD-based instance store volumes for your instance, the number of write IOPS that you can achieve decreases. This is due to the extra work the SSD controller must do to find available space, rewrite existing data, and erase unused space so that it can be rewritten. This process of garbage collection results in internal write amplification to the SSD, expressed as the ratio of SSD write operations to user write operations. This decrease in performance is even larger if the write operations are not in multiples of 4,096 bytes or not aligned to a 4,096-byte boundary. If you write a smaller amount of bytes or bytes that are not aligned, the SSD controller must read the surrounding data and store the result in a new location. This pattern results in significantly increased write amplification, increased latency, and dramatically reduced I/O performance.

SSD controllers can use several strategies to reduce the impact of write amplification. One such strategy is to reserve space in the SSD instance storage so that the controller can more efficiently manage the space available for write operations. This is called over-provisioning. The SSD-based instance store volumes provided to an instance don’t have any space reserved for over-provisioning. To reduce write amplification, we recommend that you leave 10% of the volume unpartitioned so that the SSD controller can use it for over-provisioning. This decreases the storage that you can use, but increases performance even if the disk is close to full capacity.

For instance store volumes that support TRIM, you can use the TRIM command to notify the SSD controller whenever you no longer need data that you’ve written. This provides the controller with more free space, which can reduce write amplification and increase performance. For more information, see [Instance Store Volume TRIM Support](#) (p. 925).

### Instance Features

The following is a summary of features for memory optimized instances.

<table>
<thead>
<tr>
<th></th>
<th>EBS only</th>
<th>NVMe EBS</th>
<th>Instance store</th>
<th>Placement group</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>R5</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>R5a</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>R5ad</td>
<td>No</td>
<td>Yes</td>
<td>NVME *</td>
<td>Yes</td>
</tr>
<tr>
<td>R5d</td>
<td>No</td>
<td>Yes</td>
<td>NVME *</td>
<td>Yes</td>
</tr>
<tr>
<td>u-6tb1.metal</td>
<td>Yes</td>
<td></td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
### High Availability and Reliability (X1)

X1 instances support Single Device Data Correction (SDDC +1), which detects and corrects multi-bit errors. SDDC +1 uses error checking and correction code to identify and disable a failed single DRAM device.

In addition, you can implement high availability (HA) and disaster recovery (DR) solutions to meet recovery point objective (RPO), recovery time objective (RTO), and cost requirements by leveraging Amazon CloudFormation and Recover Your Instance (p. 393).

If you run an SAP HANA production environment, you also have the option of using HANA System Replication (HSR) on X1 instances. For more information about architecting HA and DR solutions on X1 instances, see SAP HANA on the Amazon Web Services Cloud: Quick Start Reference Deployment.

### Support for vCPUs

Memory optimized instances provide a high number of vCPUs, which can cause launch issues with operating systems that have a lower vCPU limit. We strongly recommend that you use the latest AMIs when you launch memory optimized instances.

The following AMIs support launching memory optimized instances:

- Amazon Linux 2 (HVM)
- Amazon Linux AMI 2016.03 (HVM) or later
- Ubuntu Server 14.04 LTS (HVM)
- Red Hat Enterprise Linux 7.1 (HVM)
- SUSE Linux Enterprise Server 12 SP1 (HVM)
- Windows Server 2019
- Windows Server 2016
- Windows Server 2012 R2
- Windows Server 2012
- Windows Server 2008 R2 64-bit
- Windows Server 2008 SP2 64-bit

---

<table>
<thead>
<tr>
<th></th>
<th>EBS only</th>
<th>NVMe EBS</th>
<th>Instance store</th>
<th>Placement group</th>
</tr>
</thead>
<tbody>
<tr>
<td>u-9tb1.metal</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>u-12tb1.metal</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>X1</td>
<td>No</td>
<td>No</td>
<td>SSD</td>
<td>Yes</td>
</tr>
<tr>
<td>X1e</td>
<td>No</td>
<td>No</td>
<td>SSD *</td>
<td>Yes</td>
</tr>
<tr>
<td>z1d</td>
<td>No</td>
<td>Yes</td>
<td>NVME *</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* The root device volume must be an Amazon EBS volume.

For more information, see the following:

- Amazon EBS and NVMe on Windows Instances (p. 875)
- Amazon EC2 Instance Store (p. 917)
- Placement Groups (p. 710)
Release Notes

- R5 and R5d instances feature a 3.1 GHz Intel Xeon Platinum 8000 series processor.
- R5a and R5ad instances feature a 2.5 GHz AMD EPYC 7000 series processor.
- The following are requirements for high memory, R5, R5a, R5ad, R5d, and z1d instances:
  - NVMe drivers must be installed. EBS volumes are exposed as NVMe block devices (p. 875).
  - Elastic Network Adapter (ENA (p. 701)) drivers must be installed.

The following AMIs meet these requirements:
- Amazon Linux 2
- Amazon Linux AMI 2018.03
- Ubuntu 14.04 (with `linux-aws` kernel) or later
- Red Hat Enterprise Linux 7.4 or later
- SUSE Linux Enterprise Server 12 SP2 or later
- CentOS 7.4.1708 or later
- FreeBSD 11.1 or later
- R5, R5a, R5ad, R5d, and z1d instances support a maximum of 28 attachments, including network interfaces, EBS volumes, and NVMe instance store volumes. Every instance has at least one network interface attachment. For example, if you have no additional network interface attachments on an EBS-only instance, you could attach 27 EBS volumes to that instance.
- Launching a bare metal instance boots the underlying server, which includes verifying all hardware and firmware components. This means that it can take 20 minutes from the time the instance enters the running state until it becomes available over the network.
- To attach or detach EBS volumes or secondary network interfaces from a bare metal instance requires PCIe native hotplug support.
- Bare metal instances use a PCI-based serial device rather than an I/O port-based serial device. The upstream Linux kernel and the latest Amazon Linux AMIs support this device. Bare metal instances also provide an ACPI SPCR table to enable the system to automatically use the PCI-based serial device. The latest Windows AMIs automatically use the PCI-based serial device.
- You can’t launch X1 instances using a Windows Server 2008 SP2 64-bit AMI, except for x1.16xlarge instances.
- You can’t launch X1e instances using a Windows Server 2008 SP2 64-bit AMI.
- With earlier versions of the Windows Server 2008 R2 64-bit AMI, you can’t launch r4.large and r4.4xlarge instances. If you experience this issue, update to the latest version of this AMI.
- There is a limit on the total number of instances that you can launch in a region, and there are additional limits on some instance types. For more information, see How many instances can I run in Amazon EC2?. To request a limit increase, use the Amazon EC2 Instance Request Form.

Storage Optimized Instances

Storage optimized instances are designed for workloads that require high, sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications.

D2 Instances

D2 instances are well suited for the following applications:
- Massive parallel processing (MPP) data warehouse
- MapReduce and Hadoop distributed computing
• Log or data processing applications

**H1 Instances**

H1 instances are well suited for the following applications:

• Data-intensive workloads such as MapReduce and distributed file systems
• Applications requiring sequential access to large amounts of data on direct-attached instance storage
• Applications that require high-throughput access to large quantities of data

**I3 and I3en Instances**

These instances are well suited for the following applications:

• High frequency online transaction processing (OLTP) systems
• Relational databases
• NoSQL databases
• Cache for in-memory databases (for example, Redis)
• Data warehousing applications
• Distributed file systems

Bare metal instances provide your applications with direct access to physical resources of the host server, such as processors and memory. These instances are well suited for the following:

• Workloads that require access to low-level hardware features (for example, Intel VT) that are not available or fully supported in virtualized environments
• Applications that require a non-virtualized environment for licensing or support

For more information, see Amazon EC2 I3 Instances.

**Contents**

• Hardware Specifications (p. 174)
• Instance Performance (p. 175)
• Network Performance (p. 176)
• SSD I/O Performance (p. 176)
• Instance Features (p. 177)
• Release Notes (p. 178)

**Hardware Specifications**

The primary data storage for D2 instances is HDD instance store volumes. The primary data storage for I3 instances is non-volatile memory express (NVMe) SSD instance store volumes.

Instance store volumes persist only for the life of the instance. When you stop or terminate an instance, the applications and data in its instance store volumes are erased. We recommend that you regularly back up or replicate important data in your instance store volumes. For more information, see Amazon EC2 Instance Store (p. 917) and SSD Instance Store Volumes (p. 924).

The following is a summary of the hardware specifications for storage optimized instances.
<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Memory (GiB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d2.xlarge</td>
<td>4</td>
<td>30.5</td>
</tr>
<tr>
<td>d2.2xlarge</td>
<td>8</td>
<td>61</td>
</tr>
<tr>
<td>d2.4xlarge</td>
<td>16</td>
<td>122</td>
</tr>
<tr>
<td>d2.8xlarge</td>
<td>36</td>
<td>244</td>
</tr>
<tr>
<td>h1.2xlarge</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>h1.4xlarge</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>h1.8xlarge</td>
<td>32</td>
<td>128</td>
</tr>
<tr>
<td>h1.16xlarge</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>i3.large</td>
<td>2</td>
<td>15.25</td>
</tr>
<tr>
<td>i3.xlarge</td>
<td>4</td>
<td>30.5</td>
</tr>
<tr>
<td>i3.2xlarge</td>
<td>8</td>
<td>61</td>
</tr>
<tr>
<td>i3.4xlarge</td>
<td>16</td>
<td>122</td>
</tr>
<tr>
<td>i3.8xlarge</td>
<td>32</td>
<td>244</td>
</tr>
<tr>
<td>i3.16xlarge</td>
<td>64</td>
<td>488</td>
</tr>
<tr>
<td>i3.metal</td>
<td>72</td>
<td>512</td>
</tr>
<tr>
<td>i3en.xlarge</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>i3en.2xlarge</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>i3en.3xlarge</td>
<td>12</td>
<td>96</td>
</tr>
<tr>
<td>i3en.6xlarge</td>
<td>24</td>
<td>192</td>
</tr>
<tr>
<td>i3en.12xlarge</td>
<td>48</td>
<td>384</td>
</tr>
<tr>
<td>i3en.24xlarge</td>
<td>96</td>
<td>768</td>
</tr>
<tr>
<td>i3en.metal</td>
<td>96</td>
<td>768</td>
</tr>
</tbody>
</table>

For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.

For more information about specifying CPU options, see Optimizing CPU Options (p. 447).

Instance Performance

For instances with NVMe instance store volumes, be sure to use the AWS NVMe driver. For more information, see AWS NVMe Drivers for Windows Instances (p. 446).

EBS-optimized instances enable you to get consistently high performance for your EBS volumes by eliminating contention between Amazon EBS I/O and other network traffic from your instance. Some
storage optimized instances are EBS-optimized by default at no additional cost. For more information, see Amazon EBS–Optimized Instances (p. 876).

**Network Performance**

You can enable enhanced networking capabilities on supported instance types. Enhanced networking provides significantly higher packet-per-second (PPS) performance, lower network jitter, and lower latencies. For more information, see Enhanced Networking on Windows (p. 700).

Instance types that use the Elastic Network Adapter (ENA) for enhanced networking deliver high packet per second performance with consistently low latencies. Most applications do not consistently need a high level of network performance, but can benefit from having access to increased bandwidth when they send or receive data. Instance sizes that use the ENA and are documented with network performance of “Up to 10 Gbps” or “Up to 25 Gbps” use a network I/O credit mechanism to allocate network bandwidth to instances based on average bandwidth utilization. These instances accrue credits when their network bandwidth is below their baseline limits, and can use these credits when they perform network data transfers.

The following is a summary of network performance for storage optimized instances that support enhanced networking.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Network performance</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>i3.4xlarge and smaller</td>
<td>Up to 10 Gbps, use network I/O credit mechanism</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>i3.8xlarge</td>
<td>h1.8xlarge</td>
<td>10 Gbps</td>
</tr>
<tr>
<td>i3.en.3xlarge and smaller</td>
<td>Up to 25 Gbps, use network I/O credit mechanism</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>i3.16xlarge</td>
<td>i3.metal</td>
<td>i3en.6xlarge</td>
</tr>
<tr>
<td>i3en.12xlarge</td>
<td>50 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>i3en.24xlarge</td>
<td>100 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>d2.xlarge</td>
<td>Moderate</td>
<td>Intel 82599 VF (p. 707)</td>
</tr>
<tr>
<td>d2.2xlarge</td>
<td>d2.4xlarge</td>
<td>10 Gbps</td>
</tr>
</tbody>
</table>

**SSD I/O Performance**

If you use all the SSD-based instance store volumes available to your instance, you get the IOPS (4,096 byte block size) performance listed in the following table (at queue depth saturation). Otherwise, you get lower IOPS performance.

<table>
<thead>
<tr>
<th>Instance Size</th>
<th>100% Random Read IOPS</th>
<th>Write IOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>i3.large *</td>
<td>100,125</td>
<td>35,000</td>
</tr>
<tr>
<td>i3.xlarge *</td>
<td>206,250</td>
<td>70,000</td>
</tr>
<tr>
<td>i3.2xlarge</td>
<td>412,500</td>
<td>180,000</td>
</tr>
</tbody>
</table>
## Instance Size

<table>
<thead>
<tr>
<th>Instance Size</th>
<th>100% Random Read IOPS</th>
<th>Write IOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>i3.4xlarge</td>
<td>825,000</td>
<td>360,000</td>
</tr>
<tr>
<td>i3.8xlarge</td>
<td>1.65 million</td>
<td>720,000</td>
</tr>
<tr>
<td>i3.16xlarge</td>
<td>3.3 million</td>
<td>1.4 million</td>
</tr>
<tr>
<td>i3.metal</td>
<td>3.3 million</td>
<td>1.4 million</td>
</tr>
<tr>
<td>i3en.large*</td>
<td>42,500</td>
<td>32,500</td>
</tr>
<tr>
<td>i3en.xlarge*</td>
<td>85,000</td>
<td>65,000</td>
</tr>
<tr>
<td>i3en.2xlarge*</td>
<td>170,000</td>
<td>130,000</td>
</tr>
<tr>
<td>i3en.3xlarge</td>
<td>250,000</td>
<td>200,000</td>
</tr>
<tr>
<td>i3en.6xlarge</td>
<td>500,000</td>
<td>400,000</td>
</tr>
<tr>
<td>i3en.12xlarge</td>
<td>1 million</td>
<td>800,000</td>
</tr>
<tr>
<td>i3en.24xlarge</td>
<td>2 million</td>
<td>1.6 million</td>
</tr>
<tr>
<td>i3en.metal</td>
<td>2 million</td>
<td>1.6 million</td>
</tr>
</tbody>
</table>

* For these instances, you can get up to the specified performance.

As you fill your SSD-based instance store volumes, the I/O performance that you get decreases. This is due to the extra work that the SSD controller must do to find available space, rewrite existing data, and erase unused space so that it can be rewritten. This process of garbage collection results in internal write amplification to the SSD, expressed as the ratio of SSD write operations to user write operations. This decrease in performance is even larger if the write operations are not in multiples of 4,096 bytes or not aligned to a 4,096-byte boundary. If you write a smaller amount of bytes or bytes that are not aligned, the SSD controller must read the surrounding data and store the result in a new location. This pattern results in significantly increased write amplification, increased latency, and dramatically reduced I/O performance.

SSD controllers can use several strategies to reduce the impact of write amplification. One such strategy is to reserve space in the SSD instance storage so that the controller can more efficiently manage the space available for write operations. This is called *over-provisioning*. The SSD-based instance store volumes provided to an instance don’t have any space reserved for over-provisioning. To reduce write amplification, we recommend that you leave 10% of the volume unpartitioned so that the SSD controller can use it for over-provisioning. This decreases the storage that you can use, but increases performance even if the disk is close to full capacity.

For instance store volumes that support TRIM, you can use the TRIM command to notify the SSD controller whenever you no longer need data that you’ve written. This provides the controller with more free space, which can reduce write amplification and increase performance. For more information, see [Instance Store Volume TRIM Support (p. 925)](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-store-volume-trim-support.html).

## Instance Features

The following is a summary of features for storage optimized instances:

<table>
<thead>
<tr>
<th></th>
<th>EBS only</th>
<th>Instance store</th>
<th>Placement group</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2</td>
<td>No</td>
<td>HDD</td>
<td>Yes</td>
</tr>
<tr>
<td>EBS only</td>
<td>Instance store</td>
<td>Placement group</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>H1</td>
<td>No</td>
<td>HDD *</td>
<td></td>
</tr>
<tr>
<td>I3</td>
<td>No</td>
<td>NVMe *</td>
<td></td>
</tr>
<tr>
<td>I3en</td>
<td>No</td>
<td>NVMe *</td>
<td></td>
</tr>
</tbody>
</table>

*The root device volume must be an Amazon EBS volume.

For more information, see the following:

- Amazon EBS and NVMe on Windows Instances (p. 875)
- Amazon EC2 Instance Store (p. 917)
- Placement Groups (p. 710)

**Release Notes**

- You must launch storage optimized instances using an HVM AMI.
- The following are requirements for I3en and i3.metal instances:
  - NVMe drivers must be installed. EBS volumes are exposed as NVMe block devices (p. 875).
  - Elastic Network Adapter (ENA (p. 701)) drivers must be installed.

The following AMIs meet these requirements:

- Amazon Linux 2
- Amazon Linux AMI 2018.03
- Ubuntu 14.04 (with linux-aws kernel) or later
- Red Hat Enterprise Linux 7.4 or later
- SUSE Linux Enterprise Server 12 SP2 or later
- CentOS 7.4.1708 or later
- FreeBSD 11.1 or later

- Launching an i3.metal instance boots the underlying server, which includes verifying all hardware and firmware components. This means that it can take 20 minutes from the time the instance enters the running state until it becomes available over the network.
- To attach or detach EBS volumes or secondary network interfaces from a bare metal instance requires PCIe native hotplug support.
- Bare metal instances use a PCI-based serial device rather than an I/O port-based serial device. The upstream Linux kernel and the latest Amazon Linux AMIs support this device. Bare metal instances also provide an ACPI SPCR table to enable the system to automatically use the PCI-based serial device. The latest Windows AMIs automatically use the PCI-based serial device.
- There is a limit on the total number of instances that you can launch in a region, and there are additional limits on some instance types. For more information, see How many instances can I run in Amazon EC2?. To request a limit increase, use the Amazon EC2 Instance Request Form.

**Windows Accelerated Computing Instances**

If you require high processing capability, you'll benefit from using accelerated computing instances, which provide access to hardware-based compute accelerators such as Graphics Processing Units (GPUs) or Field Programmable Gate Arrays (FPGAs). Accelerated computing instances enable more parallelism for higher throughput on compute-intensive workloads.
GPU-based instances provide access to NVIDIA GPUs with thousands of compute cores. You can use GPU-based accelerated computing instances to accelerate scientific, engineering, and rendering applications by leveraging the CUDA or Open Computing Language (OpenCL) parallel computing frameworks. You can also use them for graphics applications, including game streaming, 3-D application streaming, and other graphics workloads.

If your application needs a small amount of additional graphics acceleration, but is better suited for an instance type with different compute, memory, or storage specifications, use an Elastic Graphics accelerator instead. For more information, see Amazon Elastic Graphics (p. 520).

FPGA-based instances provide access to large FPGAs with millions of parallel system logic cells. You can use FPGA-based accelerated computing instances to accelerate workloads such as genomics, financial analysis, real-time video processing, big data analysis, and security workloads by leveraging custom hardware accelerations. You can develop these accelerations using hardware description languages such as Verilog or VHDL, or by using higher-level languages such as OpenCL parallel computing frameworks. You can either develop your own hardware acceleration code or purchase hardware accelerations through the AWS Marketplace.

Important
FPGA-based instances do not support Microsoft Windows.

You can cluster accelerated computing instances into a cluster placement group. Cluster placement groups provide low latency and high-bandwidth connectivity between the instances within a single Availability Zone. For more information, see Placement Groups (p. 710).

Contents
- Accelerated Computing Instance Families (p. 179)
- Hardware Specifications (p. 181)
- Instance Performance (p. 181)
- Network Performance (p. 182)
- Instance Features (p. 182)
- Release Notes (p. 183)
- AMIs for GPU-Based Accelerated Computing Instances (p. 183)
- Installing the NVIDIA Driver on Windows (p. 184)
- Activate NVIDIA GRID Virtual Applications on G3 Instances (p. 187)
- Optimizing GPU Settings (p. 187)

For information about Linux accelerated computing instances, see Linux Accelerated Computing Instances in the Amazon EC2 User Guide for Linux Instances.

Accelerated Computing Instance Families

Accelerated computing instance families use hardware accelerators, or co-processors, to perform some functions, such as floating point number calculations, graphics processing, or data pattern matching, more efficiently than is possible in software running on CPUs. The following accelerated computing instance families are available for you to launch in Amazon EC2.

P3 Instances

P3 instances use NVIDIA Tesla V100 GPUs and are designed for general purpose GPU computing using the CUDA or OpenCL programming models or through a machine learning framework. P3 instances provide high-bandwidth networking, powerful half, single, and double-precision floating-point capabilities, and up to 32 GiB of memory per GPU, which makes them ideal for deep learning, computational fluid dynamics, computational finance, seismic analysis, molecular modeling, genomics,
rendering, and other server-side GPU compute workloads. Tesla V100 GPUs do not support graphics mode. For more information, see Amazon EC2 P3 Instances.

P3 instances support NVIDIA NVLink peer to peer transfers.

To view topology information about the system, run the following command:

```
nvidia-smi topo -m
```

For more information, see NVIDIA NVLink.

P2 Instances

P2 instances use NVIDIA Tesla K80 GPUs and are designed for general purpose GPU computing using the CUDA or OpenCL programming models. P2 instances provide high-bandwidth networking, powerful single and double precision floating-point capabilities, and 12 GiB of memory per GPU, which makes them ideal for deep learning, graph databases, high-performance databases, computational fluid dynamics, computational finance, seismic analysis, molecular modeling, genomics, rendering, and other server-side GPU compute workloads.

P2 instances support NVIDIA GPUDirect peer to peer transfers.

To view topology information about the system, run the following command:

```
nvidia-smi topo -m
```

For more information, see NVIDIA GPUDirect.

G4 Instances

G4 instances use NVIDIA Tesla GPUs and provide a cost-effective, high-performance platform for general purpose GPU computing using the CUDA or machine learning frameworks along with graphics applications using DirectX or OpenGL. G4 instances provide high-bandwidth networking, powerful half and single-precision floating-point capabilities, along with INT8 and INT4 precisions. Each GPU has 16 GiB of GDDR6 memory, making G4 instances well-suited for machine learning inference, video transcoding, and graphics applications like remote graphics workstations and game streaming in the cloud.

G4 instances support NVIDIA GRID Virtual Workstation. For more information, see NVIDIA Marketplace offerings.

G3 Instances

G3 instances use NVIDIA Tesla M60 GPUs and provide a cost-effective, high-performance platform for graphics applications using DirectX or OpenGL. G3 instances also provide NVIDIA GRID Virtual Workstation features, such as support for four monitors with resolutions up to 4096x2160, and NVIDIA GRID Virtual Applications. G3 instances are well-suited for applications such as 3D visualizations, graphics-intensive remote workstations, 3D rendering, video encoding, virtual reality, and other server-side graphics workloads requiring massively parallel processing power.

G3 instances support NVIDIA GRID Virtual Workstation and NVIDIA GRID Virtual Applications. To activate either of these features, see Activate NVIDIA GRID Virtual Applications on G3 Instances (p. 187).

G2 Instances

G2 instances use NVIDIA GRID K520 GPUs and provide a cost-effective, high-performance platform for graphics applications using DirectX or OpenGL. NVIDIA GRID GPUs also support NVIDIA’s fast capture and encode API operations. Example applications include video creation services, 3D visualizations, streaming graphics-intensive applications, and other server-side graphics workloads.
# Hardware Specifications

The following is a summary of the hardware specifications for accelerated computing instances.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Memory (GiB)</th>
<th>Accelerators</th>
</tr>
</thead>
<tbody>
<tr>
<td>p2.xlarge</td>
<td>4</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>p2.8xlarge</td>
<td>32</td>
<td>488</td>
<td>8</td>
</tr>
<tr>
<td>p2.16xlarge</td>
<td>64</td>
<td>732</td>
<td>16</td>
</tr>
<tr>
<td>p3.2xlarge</td>
<td>8</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>p3.8xlarge</td>
<td>32</td>
<td>244</td>
<td>4</td>
</tr>
<tr>
<td>p3.16xlarge</td>
<td>64</td>
<td>488</td>
<td>8</td>
</tr>
<tr>
<td>p3dn.24xlarge</td>
<td>96</td>
<td>768</td>
<td>8</td>
</tr>
<tr>
<td>g2.2xlarge</td>
<td>8</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>g2.8xlarge</td>
<td>32</td>
<td>60</td>
<td>4</td>
</tr>
<tr>
<td>g3s.xlarge</td>
<td>4</td>
<td>30.5</td>
<td>1</td>
</tr>
<tr>
<td>g3.4xlarge</td>
<td>16</td>
<td>122</td>
<td>1</td>
</tr>
<tr>
<td>g3.8xlarge</td>
<td>32</td>
<td>244</td>
<td>2</td>
</tr>
<tr>
<td>g3.16xlarge</td>
<td>64</td>
<td>488</td>
<td>4</td>
</tr>
<tr>
<td>g4dn.xlarge</td>
<td>4</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>g4dn.2xlarge</td>
<td>8</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>g4dn.4xlarge</td>
<td>16</td>
<td>64</td>
<td>1</td>
</tr>
<tr>
<td>g4dn.8xlarge</td>
<td>32</td>
<td>128</td>
<td>1</td>
</tr>
<tr>
<td>g4dn.12xlarge</td>
<td>48</td>
<td>192</td>
<td>4</td>
</tr>
<tr>
<td>g4dn.16xlarge</td>
<td>64</td>
<td>256</td>
<td>1</td>
</tr>
<tr>
<td>f1.2xlarge</td>
<td>8</td>
<td>122</td>
<td>1</td>
</tr>
<tr>
<td>f1.4xlarge</td>
<td>16</td>
<td>244</td>
<td>2</td>
</tr>
<tr>
<td>f1.16xlarge</td>
<td>64</td>
<td>976</td>
<td>8</td>
</tr>
</tbody>
</table>

For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.

For more information about specifying CPU options, see Optimizing CPU Options (p. 447).

# Instance Performance

EBS-optimized instances enable you to get consistently high performance for your EBS volumes by eliminating contention between Amazon EBS I/O and other network traffic from your instance. Some accelerated computing instances are EBS-optimized by default at no additional cost. For more information, see Amazon EBS–Optimized Instances (p. 876).
Network Performance

You can enable enhanced networking capabilities on supported instance types. Enhanced networking provides significantly higher packet-per-second (PPS) performance, lower network jitter, and lower latencies. For more information, see Enhanced Networking on Windows (p. 700).

Instance types that use the Elastic Network Adapter (ENA) for enhanced networking deliver high packet per second performance with consistently low latencies. Most applications do not consistently need a high level of network performance, but can benefit from having access to increased bandwidth when they send or receive data. Instance sizes that use the ENA and are documented with network performance of "Up to 10 Gbps" or "Up to 25 Gbps" use a network I/O credit mechanism to allocate network bandwidth to instances based on average bandwidth utilization. These instances accrue credits when their network bandwidth is below their baseline limits, and can use these credits when they perform network data transfers.

The following is a summary of network performance for accelerated computing instances that support enhanced networking.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Network performance</th>
<th>Enhanced networking</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1.2xlarge</td>
<td>Up to 10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>f1.4xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g3.4xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p3.2xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g3s.xlarge</td>
<td>10 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>g3.8xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p2.8xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p3.8xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g4dn.xlarge</td>
<td>Up to 25 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>g4dn.2xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g4dn.4xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f1.16xlarge</td>
<td>25 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>g3.16.xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p2.16xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p3.16xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g4dn.8xlarge</td>
<td>50 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
<tr>
<td>g4dn.12xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g4dn.16xlarge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p3dn.24xlarge</td>
<td>100 Gbps</td>
<td>ENA (p. 701)</td>
</tr>
</tbody>
</table>

Instance Features

The following is a summary of features for accelerated computing instances.

<table>
<thead>
<tr>
<th>EBS only</th>
<th>NVMe EBS</th>
<th>Instance store</th>
<th>Placement group</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>No</td>
<td>SSD</td>
<td>Yes</td>
</tr>
<tr>
<td>G3</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>G4</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>P2</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>P3</td>
<td>p3dn.24xlarge: No</td>
<td>p3dn.24xlarge: Yes</td>
<td>p3dn.24xlarge: NVMe *</td>
</tr>
<tr>
<td>EBS only</td>
<td>NVMe EBS</td>
<td>Instance store</td>
<td>Placement group</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>F1</td>
<td>No</td>
<td>No</td>
<td>NVMe *</td>
</tr>
</tbody>
</table>

* The root device volume must be an Amazon EBS volume.

For more information, see the following:

- Amazon EBS and NVMe on Windows Instances (p. 875)
- Amazon EC2 Instance Store (p. 917)
- Placement Groups (p. 710)

### Release Notes

- You must launch the instance using an HVM AMI.
- The following are requirements for G4 instances:
  - NVMe drivers must be installed. EBS volumes are exposed as NVMe block devices (p. 875).
  - Elastic Network Adapter (ENA (p. 701)) drivers must be installed.

The following AMIs meet these requirements:

- Amazon Linux 2
- Amazon Linux AMI 2018.03
- Ubuntu 14.04 (with linux-aws kernel) or later
- Red Hat Enterprise Linux 7.4 or later
- SUSE Linux Enterprise Server 12 SP2 or later
- CentOS 7.4.1708 or later
- FreeBSD 11.1 or later
- GPU-based instances can’t access the GPU unless the NVIDIA drivers are installed. For more information, see Installing the NVIDIA Driver on Windows (p. 184).
- There is a limit of 100 AFIs per region.
- There is a limit on the number of instances that you can run. For more information, see How many instances can I run in Amazon EC2? in the Amazon EC2 FAQ. To request an increase in these limits, use the following form: Request to Increase Amazon EC2 Instance Limit.
- If you launch a multi-GPU instance with a Windows AMI that was created on a single-GPU instance, Windows does not automatically install the NVIDIA driver for all GPUs. You must authorize the driver installation for the new GPU hardware. You can correct this manually in the Device Manager by opening the Other device category (the inactive GPUs do not appear under Display Adapters). For each inactive GPU, open the context (right-click) menu, choose Update Driver Software, and then choose the default Automatic Update option.
- When using Microsoft Remote Desktop Protocol (RDP), GPUs that use the WDDM driver model are replaced with a non-accelerated Remote Desktop display driver. To access your GPU hardware, you must use a different remote access tool, such as Teradici Cloud Access Software, NICE Desktop Cloud Visualization (DCV), or VNC. You can also use one of the GPU AMIs from the AWS Marketplace because they provide remote access tools that support 3D acceleration.

### AMIs for GPU-Based Accelerated Computing Instances

To help you get started, NVIDIA and others provide AMIs for GPU-based accelerated computing instances. These reference AMIs include the NVIDIA driver, which enables full functionality and performance of the NVIDIA GPUs.
For a list of AMIs with the NVIDIA driver, search AWS Marketplace as follows:

- NVIDIA P3 AMIs
- NVIDIA Quadro Virtual Workstation P3 AMIs
- NVIDIA P2 AMIs
- NVIDIA GRID G4 AMIs
- NVIDIA GRID G3 AMIs
- NVIDIA GRID G2 AMIs

You can launch accelerated computing instances using any HVM AMI.

Important
These AMIs include drivers, software, or toolkits that are developed, owned, or provided by NVIDIA Corporation. By using these AMIs, you agree to use these NVIDIA drivers, software, or toolkits only on Amazon EC2 instances that include NVIDIA hardware.

You can also install the NVIDIA driver. For more information, see Installing the NVIDIA Driver on Windows (p. 184).

Installing the NVIDIA Driver on Windows

A GPU-based accelerated computing instance must have the appropriate NVIDIA driver. Depending on the instance type, you can either download a public NVIDIA driver, use an NVIDIA Marketplace offering, or download a driver from Amazon S3 that is available only to AWS customers.

Contents
- Public NVIDIA Drivers (p. 184)
- NVIDIA GRID Drivers for G4 Instances (p. 185)
- NVIDIA GRID Drivers for G3 Instances (p. 186)

Public NVIDIA Drivers

To install a public NVIDIA driver on your Windows instance, log on to your instance as the administrator using Remote Desktop and complete the following procedure.

To install a public NVIDIA driver

1. Download the public NVIDIA driver that is appropriate for your instance type from http://www.nvidia.com/Download/Find.aspx.

<table>
<thead>
<tr>
<th>Instances</th>
<th>Product Type</th>
<th>Product Series</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>G4 †</td>
<td>Tesla</td>
<td>T-Series</td>
<td>Tesla T4</td>
</tr>
<tr>
<td>G3</td>
<td>Tesla</td>
<td>M-Class</td>
<td>M60</td>
</tr>
<tr>
<td>G2</td>
<td>GRID</td>
<td>GRID Series</td>
<td>GRID K520</td>
</tr>
<tr>
<td>P3</td>
<td>Tesla</td>
<td>V-Series</td>
<td>V100</td>
</tr>
<tr>
<td>P2</td>
<td>Tesla</td>
<td>K-Series</td>
<td>K-80</td>
</tr>
</tbody>
</table>

† G4 instances require driver version 426.00 or later.
2. Open the folder where you downloaded the driver and launch the installation file. Follow the instructions to install the driver and reboot your instance as required.

3. Disable the built-in display adapter using Device Manager. Install the Windows features Media Foundation and Quality Windows Audio Video Experience.

4. To verify that the GPU is working correctly, check Device Manager.

5. Complete the optimization steps in Optimizing GPU Settings (p. 187) to achieve the best performance from your GPU.

When you use P3 instances for graphics applications, you can use AMIs published by NVIDIA that enable NVIDIA Quadro Virtual Workstation. For G4 instances, see NVIDIA GRID Drivers for G4 Instances (p. 185).

NVIDIA GRID Drivers for G4 Instances

There are two ways that you can use NVIDIA GRID software for graphics applications on G4 instances. You can use AMIs with GRID preinstalled or download the NVIDIA GRID vGaming driver from Amazon S3 and install it on your G4 instances.

If you own GRID licenses, you should be able to use those licenses on your G4 instances. For more information, see NVIDIA GRID Software Quick Start Guide.

Option 1: Use an AMI with GRID for your G4 instances

There are two types of AMIs available.

- NVIDIA Marketplace offerings featuring NVIDIA Quadro Virtual Workstation software. These AMIs support running up to four 4K displays per GPU on a G4 instance. To find these AMIs, use this link: NVIDIA Marketplace offerings.
- NVIDIA Marketplace offerings featuring NVIDIA Gaming drivers. These AMIs support running a single 4K display per GPU on a G4 instance. To find these AMIs, use this link: NVIDIA Marketplace offerings.

Option 2: Use the NVIDIA GRID vGaming driver

This driver is available to AWS customers only. By downloading, you agree to use the downloaded software only to develop AMIs for use with the NVIDIA Tesla T4 hardware. Upon installation of the software, you are bound by the terms of the NVIDIA GRID Cloud End User License Agreement.

If you own GRID licenses, you should be able to use those licenses on your G4 instances. For more information, see NVIDIA GRID Software Quick Start Guide.

Use the following procedure to install this driver.

1. Connect to your Windows instance.

2. Download and install the NVIDIA GRID driver from Amazon S3 using this link: NVIDIA Windows Gaming Driver for G4 Instances.

3. Use the following steps to create the required configuration.
   a. Open the registry editor.
   b. In the left pane, navigate to HKEY_LOCAL_MACHINE\SOFTWARE\NVIDIA Corporation \Global.
   c. In the right pane, right-click and choose New, DWORD (32-bit) value.
   d. Set the name to \vGamingMarketplace and the value to 2.

4. Download the certification file, rename the file GridSwCert.txt, and move the file to the Public Documents folder on your system drive. Typically, the folder path is C:\Users\Public\Public Documents (Windows Explorer) or C:\Users\Public\Documents (Command Prompt window).
5. Reboot your instance.
6. Verify the vGaming license using the following command.

```
"C:\Program Files\NVIDIA Corporation\NVS Mi\nvidia-smi.exe" -q
```

The output should be similar to the following.

<table>
<thead>
<tr>
<th>GRID Licensed Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Name</td>
</tr>
<tr>
<td>License Status</td>
</tr>
</tbody>
</table>

**NVIDIA GRID Drivers for G3 Instances**

AWS provides Windows AMIs for G3 instances that come with GRID drivers preinstalled. To find these AMIs, open NVIDIA Marketplace offerings.

Alternatively, you can download the NVIDIA GRID driver from Amazon S3 using the AWS Tools for Windows PowerShell as described in the following procedure.

**Important**

This download is available to AWS customers only. By downloading, you agree that you will use the downloaded software only to develop AMIs for use with the NVIDIA Tesla M60 hardware. Upon installation of the software, you are bound by the terms of the NVIDIA GRID Cloud End User License Agreement.

**Prerequisites**

- Confirm that the version of Windows you are launching is supported by NVIDIA GRID drivers. Driver requirements are documented at https://docs.nvidia.com/grid/index.html.
- Associate an IAM role with your instance that has permissions to use the `s3:ListBucket` and `s3:GetObject` actions. For more information, see IAM Roles for Amazon EC2 (p. 646). Alternatively, configure the Tools for Windows PowerShell to use your AWS credentials. For more information, see Using AWS Credentials.

**To install the NVIDIA GRID driver on a G3 instance**

1. Open a PowerShell window.
2. Download the drivers and the NVIDIA GRID Cloud End User License Agreement to your desktop with the following PowerShell commands (you can copy and paste the entire block of commands at one time).

```powershell
$Bucket = "ec2-windows-nvidia-drivers"
$KeyPrefix = "latest"
$LocalPath = "C:\Users\Administrator\Desktop\NVIDIA"
$Objects = Get-S3Object -BucketName $Bucket -KeyPrefix $KeyPrefix -Region us-east-1
foreach ($Object in $Objects) {
    $LocalFileName = $Object.Key
    if ($LocalFileName -ne "") -and $Object.Size -ne 0) {
        $LocalFilePath = Join-Path $LocalPath $LocalFileName
        Copy-S3Object -BucketName $Bucket -Key $Object.Key -LocalFile $LocalFilePath -Region us-east-1
    }
}
```

Multiple versions of the NVIDIA GRID driver are stored in this bucket. You can download all of the available versions in the bucket by removing the `KeyPrefix $KeyPrefix` option.
3. Navigate to the desktop and double-click the installation file to launch it (choose the driver version that corresponds to your instance OS version). Follow the instructions to install the driver and reboot your instance as required. To verify that the GPU is working properly, check Device Manager.

4. Disable the licensing page in the control panel to prevent users from accidentally changing the product type (NVIDIA GRID Virtual Workstation is enabled by default). For more information, see the GRID Licensing User Guide.
   a. Run `regedit.exe` to open the registry editor.
   b. Navigate to `HKEY_LOCAL_MACHINE\SOFTWARE\NVIDIA Corporation\Global\GridLicensing`.
   c. Open the context (right-click) menu on the right pane and choose `New, DWORD`.
   d. For `Name`, enter `NvCplDisableManageLicensePage` and type `Enter`.
   e. Open the context (right-click) menu on `NvCplDisableManageLicensePage` and choose `Modify`.
   f. For `Value data`, type `1` and choose `OK`.

5. (Optional) To enable NVIDIA GRID Virtual Applications, complete the GRID activation steps in Activate NVIDIA GRID Virtual Applications on G3 Instances (p. 187) (NVIDIA GRID Virtual Workstation is enabled by default).

6. Complete the optimization steps in Optimizing GPU Settings (p. 187) to achieve the best performance from your GPU.

**Activate NVIDIA GRID Virtual Applications on G3 Instances**

To activate the GRID Virtual Applications on G3 instances (NVIDIA GRID Virtual Workstation is enabled by default), you must define the product type for the driver in the registry.

**To activate GRID Virtual Applications on G3 Windows instances**

1. Run `regedit.exe` to open the registry editor.
2. Navigate to `HKEY_LOCAL_MACHINE\SOFTWARE\NVIDIA Corporation\Global\GridLicensing`.
3. Open the context (right-click) menu on the right pane and choose `New, DWORD`.
4. For `Name`, enter `FeatureType` and type `Enter`.
5. Open the context (right-click) menu on `FeatureType` and choose `Modify`.
6. For `Value data`, enter `0` for NVIDIA GRID Virtual Applications and choose `OK`.
7. Open the context (right-click) menu on the right pane and choose `New, DWORD`.
8. For `Name`, enter `IgnoreSP` and type `Enter`.
9. Open the context (right-click) menu on `IgnoreSP` and choose `Modify`.
10. For `Value data`, type `1` and choose `OK`.
11. Close the registry editor.

**Optimizing GPU Settings**

There are several GPU setting optimizations that you can perform to achieve the best performance on G3, G4, P2, P3, and P3dn instances. By default, the NVIDIA driver uses an autoboost feature, which varies the GPU clock speeds. By disabling the autoboost feature and setting the GPU clock speeds to their maximum frequency, you can consistently achieve the maximum performance with your GPU instances.

**To optimize GPU settings**

1. Open a PowerShell window and navigate to the NVIDIA installation folder.
cd "C:\Program Files\NVIDIA Corporation\NVSMI"

2. Disable the autoboost feature for all GPUs on the instance.

`.\nvidia-smi --auto-boost-default=0`

**Note**
GPUs on P3, P3dn, and G4 instances do not support autoboost.

3. Set all GPU clock speeds to their maximum frequency. Use the memory and graphics clock speeds specified in the following commands.

**Note**
Some versions of the NVIDIA driver do not allow setting application clock speed and throw a "Setting applications clocks is not supported for GPU ..." error, which you can ignore.

- **G3 instances:**
  `.\nvidia-smi -ac "2505,1177"`

- **G4 instances:**
  `.\nvidia-smi -ac "5001,1590"`

- **P2 instances:**
  `.\nvidia-smi -ac "2505,875"`

- **P3 and P3dn instances:**
  `.\nvidia-smi -ac "877,1530"

## Changing the Instance Type

As your needs change, you might find that your instance is over-utilized (the instance type is too small) or under-utilized (the instance type is too large). If this is the case, you can change the size of your instance. For example, if your `t2.micro` instance is too small for its workload, you can change it to another instance type that is appropriate for the workload.

You might also want to migrate from a previous generation instance type to a current generation instance type to take advantage of some features; for example, support for IPv6.

You can change the size of the instance simply by changing its instance type, which is known as **resizing** it.

When you resize an instance, you must select an instance type that is compatible with the configuration of the instance. If the instance type that you want is not compatible with the instance configuration you have, then you must migrate your application to a new instance with the instance type that you need.

**Important**
When you resize an instance, the resized instance usually has the same number of instance store volumes that you specified when you launched the original instance. With instance types that support NVMe instance store volumes (which are available by default), the resized instance might have additional instance store volumes, depending on the AMI. Otherwise, you can
migrate your application to an instance with a new instance type manually, specifying the number of instance store volumes that you need when you launch the new instance.

Contents
- Compatibility for Resizing Instances (p. 189)
- Resizing an Amazon EBS–backed Instance (p. 189)
- Migrating to a New Instance Configuration (p. 191)

Compatibility for Resizing Instances

You can resize an instance only if its current instance type and the new instance type that you want are compatible in the following ways:

- **Architecture:** AMIs are specific to the architecture of the processor, so you must select an instance type with the same processor architecture as the current instance type. For example:
  - A1 instances are the only instances that support processors based on the Arm architecture. If you are resizing an instance type with a processor based on the Arm architecture, you are limited to the instance types that support a processor based on the Arm architecture.
  - The following instance types are the only instance types that support 32-bit AMIs: t2.nano, t2.micro, t2.small, t2.medium, c3.large, t1.micro, m1.small, m1.medium, and c1.medium. If you are resizing a 32-bit instance, you are limited to these instance types.

- **Network:** Newer instance types must be launched in a VPC. Therefore, you can't resize an instance in the EC2-Classic platform to a instance type that is available only in a VPC unless you have a nondefault VPC. To check whether your instance is in a VPC, check the VPC ID value on the details pane of the Instances screen in the Amazon EC2 console. For more information, see Migrating from a Windows Instance in EC2-Classic to a Windows Instance in a VPC (p. 772).

- **Enhanced networking:** Instance types that support enhanced networking (p. 700) require the necessary drivers installed. For example, the C5, C5d, C5n, F1, G3, G4, H1, I3, I3en, m4.16xlarge, M5, M5a, M5ad, M5d P2, P3, R4, R5, R5a, R5ad, R5d, T3, T3a, u-6tb1.meta, u-9tb1.meta, u-12tb1.meta, X1, X1e, and z1d instance types require EBS-backed AMIs with the Elastic Network Adapter (ENA) drivers installed. To resize an existing instance to an instance type that supports enhanced networking, you must first install the ENA drivers (p. 701) or ixgbevf drivers (p. 707) on your instance, as appropriate.

- **NVMe:** EBS volumes are exposed as NVMe block devices on Nitro-based instances (p. 122). If you resize an instance from an instance type that does not support NVMe to an instance type that supports NVMe, you must first install the NVMe drivers (p. 875) on your instance. Also, the device names for devices that you specify in the block device mapping are renamed using NVMe device names (/dev/nvme[0–26]n1).

- **AMI:** For information about the AMIs required by instance types that support enhanced networking and NVMe, see the Release Notes in the following documentation:
  - General Purpose Instances (p. 125)
  - Compute Optimized Instances (p. 161)
  - Memory Optimized Instances (p. 165)
  - Storage Optimized Instances (p. 173)

Resizing an Amazon EBS–backed Instance

You must stop your Amazon EBS–backed instance before you can change its instance type. When you stop and start an instance, be aware of the following:

- We move the instance to new hardware; however, the instance ID does not change.
• If your instance has a public IPv4 address, we release the address and give it a new public IPv4 address. The instance retains its private IPv4 addresses, any Elastic IP addresses, and any IPv6 addresses.

• If your instance is in an Auto Scaling group, the Amazon EC2 Auto Scaling service marks the stopped instance as unhealthy, and may terminate it and launch a replacement instance. To prevent this, you can suspend the scaling processes for the group while you're resizing your instance. For more information, see Suspending and Resuming Scaling Processes in the Amazon EC2 Auto Scaling User Guide.

• If your instance is in a cluster placement group (p. 710) and, after changing the instance type, the instance start fails, try the following: stop all the instances in the cluster placement group, change the instance type for the affected instance, and then restart all the instances in the cluster placement group.

• Ensure that you plan for downtime while your instance is stopped. Stopping and resizing an instance may take a few minutes, and restarting your instance may take a variable amount of time depending on your application’s startup scripts.

For more information, see Stop and Start Your Instance (p. 383).

Use the following procedure to resize an Amazon EBS–backed instance using the AWS Management Console.

To resize an Amazon EBS–backed instance

1. (Optional) If the new instance type requires drivers that are not installed on the existing instance, you must connect to your instance and install the drivers first. For more information, see Compatibility for Resizing Instances (p. 189).

   Note
   The AWS PV driver package should be updated before changing instance families. For more information, see Upgrading PV Drivers on Your Windows Instances (p. 436).

2. (Optional) If you've configured your instance to use static IP addressing (p. 468) and you resize the instance from a type that doesn't support enhanced networking to an instance type that does support enhanced networking you might get a warning about a potential IP address conflict when you reconfigure static IP addressing. To prevent this, enable DHCP on the network interface for your instance before you change the instance type. From your instance, open the Network and Sharing Center, go to Internet Protocol Version 4 (TCP/IPv4) Properties for the network interface, and choose Obtain an IP address automatically. Change the instance type and reconfigure static IP addressing on the network interface.

3. Open the Amazon EC2 console.
4. [Windows Server 2016 and later] Connect to your Windows instance and run the following EC2Launch PowerShell script to configure the instance after it is resized.

   ```
   PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeInstance.ps1 - Schedule
   ```

5. In the navigation pane, choose Instances.
6. Select the instance and choose Actions, Instance State, Stop.
7. In the confirmation dialog box, choose Yes, Stop. It can take a few minutes for the instance to stop.
8. With the instance still selected, choose Actions, Instance Settings, Change Instance Type. This action is disabled if the instance state is not stopped.
9. In the Change Instance Type dialog box, do the following:
   a. From Instance Type, select the instance type that you want. If the instance type that you want does not appear in the list, then it is not compatible with the configuration of your instance (for example, because of virtualization type). For more information, see Compatibility for Resizing Instances (p. 189).
b. (Optional) If the instance type that you selected supports EBS-optimization, select **EBS-optimized** to enable EBS-optimization or deselect **EBS-optimized** to disable EBS-optimization. If the instance type that you selected is EBS-optimized by default, **EBS-optimized** is selected and you can't deselect it.

c. Choose **Apply** to accept the new settings.

10. To restart the stopped instance, select the instance and choose **Actions, Instance State, Start**.

11. In the confirmation dialog box, choose **Yes, Start**. It can take a few minutes for the instance to enter the running state.

## Migrating to a New Instance Configuration

If the current configuration of your instance is incompatible with the new instance type that you want, then you can't resize the instance to that instance type. Instead, you can migrate your application to a new instance with a configuration that is compatible with the new instance type that you want.

**To migrate your application to a compatible instance**

1. Back up any data on your instance store volumes that you need to keep to persistent storage. To migrate data on your EBS volumes that you need to keep, create a snapshot of the volumes (see Creating Amazon EBS Snapshots (p. 822)) or detach the volume from the instance so that you can attach it to the new instance later (see Detaching an Amazon EBS Volume from an Instance (p. 816)).

2. Launch a new instance, selecting the following:

   - If you are using an Elastic IP address, select the VPC that the original instance is currently running in.
   - Any EBS volumes that you detached from the original instance and want to attach to the new instance, or new EBS volumes based on the snapshots that you created.
   - If you want to allow the same traffic to reach the new instance, select the security group that is associated with the original instance.

3. Install your application and any required software on the instance.

4. Restore any data that you backed up from the instance store volumes of the original instance.

5. If you are using an Elastic IP address, assign it to the newly launched instance as follows:

   a. In the navigation pane, choose **Elastic IPs**.
   b. Select the Elastic IP address that is associated with the original instance and choose **Actions, Disassociate address**. When prompted for confirmation, choose **Disassociate address**.
   c. With the Elastic IP address still selected, choose **Actions, Associate address**.
   d. From **Instance**, select the new instance, and then choose **Associate**.

6. (Optional) You can terminate the original instance if it's no longer needed. Select the instance and verify that you are about to terminate the original instance, not the new instance (for example, check the name or launch time). Choose **Actions, Instance State, Terminate**.

## Instance Purchasing Options

Amazon EC2 provides the following purchasing options to enable you to optimize your costs based on your needs:

- **On-Demand Instances** – Pay, by the hour, for the instances that you launch.
- **Reserved Instances** – Purchase, at a significant discount, instances that are always available, for a term from one to three years.
• **Scheduled Instances** – Purchase instances that are always available on the specified recurring schedule, for a one-year term.

• **Spot Instances** – Request unused EC2 instances, which can lower your Amazon EC2 costs significantly.

• **Dedicated Hosts** – Pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs.

• **Dedicated Instances** – Pay, by the hour, for instances that run on single-tenant hardware.

• **Capacity Reservations** – Reserve capacity for your EC2 instances in a specific Availability Zone for any duration.

If you require a capacity reservation, purchase Reserved Instances or Capacity Reservations for a specific Availability Zone, or purchase Scheduled Instances. Spot Instances are a cost-effective choice if you can be flexible about when your applications run and if they can be interrupted. Dedicated Hosts or Dedicated Instances can help you address compliance requirements and reduce costs by using your existing server-bound software licenses. For more information, see Amazon EC2 Pricing.

### Contents

- Determining the Instance Lifecycle (p. 192)
- On-Demand Instances (p. 193)
- Reserved Instances (p. 197)
- Scheduled Reserved Instances (p. 227)
- Spot Instances (p. 231)
- Dedicated Hosts (p. 294)
- Dedicated Instances (p. 312)
- On-Demand Capacity Reservations (p. 317)

### Determining the Instance Lifecycle

The lifecycle of an instance starts when it is launched and ends when it is terminated. The purchasing option that you choose affects the lifecycle of the instance. For example, an On-Demand Instance runs when you launch it and ends when you terminate it. A Spot Instance runs as long as capacity is available and your maximum price is higher than the Spot price. You can launch a Scheduled Instance during its scheduled time period; Amazon EC2 launches the instances and then terminates them three minutes before the time period ends.

Use the following procedure to determine the lifecycle of an instance.

**To determine the instance lifecycle using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**.
3. Select the instance.
4. On the **Description** tab, find **Tenancy**. If the value is **host**, the instance is running on a Dedicated Host. If the value is **dedicated**, the instance is a Dedicated Instance.
5. On the **Description** tab, find **Lifecycle**. If the value is **spot**, the instance is a Spot Instance. If the value is **scheduled**, the instance is a Scheduled Instance. If the value is **normal**, the instance is either an On-Demand Instance or a Reserved Instance.
6. (Optional) If you have purchased a Reserved Instance and want to verify that it is being applied, you can check the usage reports for Amazon EC2. For more information, see Amazon EC2 Usage Reports (p. 976).
To determine the instance lifecycle using the AWS CLI

Use the following describe-instances command:

```
aws ec2 describe-instances --instance-ids i-1234567890abcdef0
```

If the instance is running on a Dedicated Host, the output contains the following information:

```
"Tenancy": "host"
```

If the instance is a Dedicated Instance, the output contains the following information:

```
"Tenancy": "dedicated"
```

If the instance is a Spot Instance, the output contains the following information:

```
"InstanceLifecycle": "spot"
```

If the instance is a Scheduled Instance, the output contains the following information:

```
"InstanceLifecycle": "scheduled"
```

Otherwise, the output does not contain InstanceLifecycle.

On-Demand Instances

An On-Demand Instance is an instance that you use on demand. You have full control over its lifecycle—you decide when to launch, stop, hibernate, start, reboot, or terminate it.

There is no long-term commitment required when you purchase On-Demand Instances. You pay only for the hours that your On-Demand Instances are in the running state. The price per hour for a running On-Demand Instance is fixed, and is listed on the On-Demand Pricing page.

We recommend that you use On-Demand Instances for applications with short-term, irregular workloads that cannot be interrupted.

For significant discounts compared to On-Demand Instances, you can use Spot Instances (p. 231) or Reserved Instances (p. 197).

Working With On-Demand Instances

You can work with On-Demand Instances in the following ways:

- Launch Your Instance (p. 333)
- Stop and Start Your Instance (p. 383)
- Hibernate Your Instance (p. 385)
- Reboot Your Instance (p. 385)
- Instance Retirement (p. 386)
- Terminate Your Instance (p. 388)
- Recover Your Instance (p. 393)
• Identify EC2 Windows Instances (p. 518)

If you're new to Amazon EC2, see How to Get Started with Amazon EC2 (p. 1).

On-Demand Instance Limits

There is a limit to the number of running On-Demand Instances per AWS account per Region. You can determine whether your On-Demand Instance limits are count-based or vCPU-based.

Note
Amazon EC2 is transitioning On-Demand Instance limits from count-based instance limits to vCPU-based instance limits. From October 24, 2019, On-Demand Instance limits will be vCPU-based only. For more information, see EC2 On-Demand Instance limits.

Count-Based Instance Limits

With count-based instance limits, your On-Demand Instance limits are managed in terms of the number of instances per instance type that you are running. For more information, see How many instances can I run in Amazon EC2? From October 24, 2019, count-based instance limits for On-Demand Instances are not available.

vCPU-based Instance Limits

With vCPU-based instance limits, your limits are managed in terms of the number of virtual central processing units (vCPUs) that your running On-Demand Instances are using, regardless of the instance type. For more information, see EC2 On-Demand Instance limits.

There are five vCPU-based On-Demand Instance limits, listed in the following table. Each limit specifies the vCPU limit for one or more instance families. For information about the different instance families, generations, and sizes, see Amazon EC2 Instance Types.

<table>
<thead>
<tr>
<th>On-Demand Instance limit name</th>
<th>Default vCPU limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running On-Demand Standard (A, C, D, H, I, M, R, T, and Z) instances</td>
<td>1152 vCPUs</td>
</tr>
<tr>
<td>Running On-Demand F instances</td>
<td>128 vCPUs</td>
</tr>
<tr>
<td>Running On-Demand G instances</td>
<td>128 vCPUs</td>
</tr>
<tr>
<td>Running On-Demand P instances</td>
<td>128 vCPUs</td>
</tr>
<tr>
<td>Running On-Demand X instances</td>
<td>128 vCPUs</td>
</tr>
</tbody>
</table>

With vCPU limits, you can use your limit in terms of the number of vCPUs required to launch any combination of instance types that meet your changing application needs. For example, with a Standard instance limit of 256 vCPUs, you could launch 32 m5.2xlarge instances (32 x 8 vCPUs) or 16 c5.4xlarge instances (16 x 16 vCPUs), or a combination of any Standard instance types and sizes that total 256 vCPUs. For more information, see EC2 On-Demand Instance limits.

Calculating How Many vCPUs You Need

You can use the vCPU limits calculator to determine the number of vCPUs that you require for your application needs.

The following screenshot shows the vCPU limits calculator.
You can view and use the following controls and information:

- **Instance type** – The instance types that you add to the vCPU limits calculator.
- **Instance count** – The number of instances that you require for the selected instance type.
- **vCPU count** – The number of vCPUs that correspond to the **Instance count**.
- **Current limit** – Your current limit for the limit type to which the instance type belongs. The limit applies to all instance types of the same limit type. For example, in the preceding screenshot, the current limit for m5.2xlarge and c5.4xlarge is 1,920 vCPUs, which is the limit for all the instance types that belong to the All Standard instances limit.
- **New limit** – The new limit is calculated by adding the values for **Current limit** and **vCPUs needed**.
- **X** – Choose the X to remove the row.
- **Add instance type** – Choose **Add instance type** to add another instance type to the calculator.
- **Limits calculation** – Displays the current limit, vCPUs needed, and new limit for the limit types.
  - **Instance limit name** – The limit type for the instance types that you selected.
  - **Current limit** – The current limit for the limit type.
  - **vCPUs needed** – The number of vCPUs that correspond to the number of instances that you specified. For the All Standard instances limit type, the vCPUs needed is calculated by adding the values for **vCPU count** for all the instance types of this limit type.
  - **New limit** – The new limit is calculated by adding the values for **Current limit** and **vCPUs needed**.
  - **Options** – Choose **Request limit increase** to request a limit increase for the corresponding limit type.

**To calculate the number of required vCPUs**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. From the navigation bar, select a Region.
3. From the left navigator, choose **Limits**.
4. Choose **Calculate vCPU limit**.
5. Choose **Add instance type**, choose the required instance type, and specify the required number of instances. To add more instance types, choose **Add instance type** again.
6. View **Limits calculation** for the required new limit.
7. When you’ve finished using the calculator, choose **Close**.

**Requesting a Limit Increase**

You can request a limit increase for each On-Demand Instance limit type from the **Limits page** or the vCPU limits calculator in the Amazon EC2 console. Complete the required fields on the AWS Support Center limit increase form with your use case. For **Primary Instance Type**, select the limit type that corresponds to the **Instance limit name** in the vCPU limits calculator. For the new limit value, use the value that appears in the **New limit** column in the vCPU limits calculator. For more information about requesting a limit increase, see Amazon EC2 Service Limits (p. 974).

**Opt-in Period for vCPU-Based Instance Limits**

The opt-in period for the new vCPU-based instance limits lasts until the deadline date, October 24, 2019. During the opt-in period, you can switch between vCPU-based and count-based instance limits. When you opt in, EC2 automatically computes your new limits in vCPUs, giving you access to launch at least the same number of instances (if not more) that you currently can launch.

After October 24, 2019, On-Demand Instance limits will be vCPU-based only; count-based limits will not be available. For more information, see EC2 On-Demand Instance limits.

**Note**

You opt in or out of vCPU-based instance limits at an AWS account level, so your limit preference applies to all Regions.

**To opt in to vCPU-based instance limits**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. From the navigation pane, choose **Limits**.
3. To opt in to vCPU instance limits, choose **Opt in**, enter **Opt in** in the confirmation window, and then choose **Opt in**.

**Note**

It can take up to 15 minutes to switch your account from count-based to vCPU-based instance limits, or from vCPU-based to count-based instance limits.

**Monitoring On-Demand Instance Limits and Usage**

You can view and manage your On-Demand Instance limits from the **Limits page** in the Amazon EC2 console and from the Amazon EC2 **AWS services page** in the Service Quotas console. For more information, see Amazon EC2 Service Limits (p. 974) in the Amazon EC2 User Guide for Linux Instances and Viewing a Service Quota in the Service Quotas User Guide.

With Amazon CloudWatch metrics integration, you can monitor EC2 usage against limits. You can also configure alarms to warn about approaching limits. For more information, see Using Amazon CloudWatch Alarms in the Service Quotas User Guide.

**Note**

If you have opted in to use vCPU-based instance limits, Trusted Advisor currently does not display EC2 On-Demand Instance limits or usage information. From October 24, 2019, Trusted Advisor will display this information. For more information, see Why don't I see a limit amount...
Reserved Instances

Reserved Instances provide you with a significant discount compared to On-Demand Instance pricing. Reserved Instances are not physical instances, but rather a billing discount applied to the use of On-Demand Instances in your account. These On-Demand Instances must match certain attributes in order to benefit from the billing discount.

The following diagram shows a basic overview of purchasing and using Reserved Instances.

In this scenario, you have a running On-Demand Instance (T2) in your account, for which you're currently paying On-Demand rates. You purchase a Reserved Instance that matches the attributes of your running instance, and the billing benefit is immediately applied. Next, you purchase a Reserved Instance for a C4 instance. You do not have any running instances in your account that match the attributes of this Reserved Instance. In the final step, you launch an instance that matches the attributes of the C4 Reserved Instance, and the billing benefit is immediately applied.

Key Variables That Determine Reserved Instance Pricing

The Reserved Instance pricing is determined by the following key variables.

Instance Attributes

A Reserved Instance has four instance attributes that determine its price. The attributes also determine how the Reserved Instance is applied to a running instance in your account.

- **Instance type**: For example, m4.large. This is composed of the instance family (m4) and the instance size (large).
- **Scope**: Whether the Reserved Instance applies to a Region (regional Reserved Instance) or specific Availability Zone (zonal Reserved Instance). For more information, see Regional and Zonal Reserved Instances (Scope) (p. 199).
• **Tenancy**: Whether your instance runs on shared (default) or single-tenant (dedicated) hardware. For more information, see [Dedicated Instances](#).

• **Platform**: The operating system; for example, Windows or Linux/Unix. For more information, see [Choosing a Platform](#).

Reserved Instances do not renew automatically; when they expire, you can continue using the EC2 instance without interruption, but you are charged On-Demand rates. In the above example, when the Reserved Instances that cover the T2 and C4 instances expire, you go back to paying the On-Demand rates until you terminate the instances or purchase new Reserved Instances that match the instance attributes.

### Term Commitment

You can purchase a Reserved Instance for a one-year or three-year commitment, with the three-year commitment offering a bigger discount.

- **One-year**: A year is defined as 31536000 seconds (365 days).
- **Three-year**: Three years is defined as 94608000 seconds (1095 days).

### Payment Options

The following payment options are available for Reserved Instances:

- **All Upfront**: Full payment is made at the start of the term, with no other costs or additional hourly charges incurred for the remainder of the term, regardless of hours used.
- **Partial Upfront**: A portion of the cost must be paid upfront and the remaining hours in the term are billed at a discounted hourly rate, regardless of whether the Reserved Instance is being used.
- **No Upfront**: You are billed a discounted hourly rate for every hour within the term, regardless of whether the Reserved Instance is being used. No upfront payment is required.

**Note**

No Upfront Reserved Instances are based on a contractual obligation to pay monthly for the entire term of the reservation. For this reason, a successful billing history is required before you can purchase No Upfront Reserved Instances.

Generally speaking, you can save more money making a higher upfront payment for Reserved Instances. You can also find Reserved Instances offered by third-party sellers at lower prices and shorter term lengths on the Reserved Instance Marketplace. For more information, see [Reserved Instance Marketplace](#).

### Offering Class

If your computing needs change, you may be able to modify or exchange your Reserved Instance, depending on the offering class.

- **Standard**: These provide the most significant discount, but can only be modified.
- **Convertible**: These provide a lower discount than Standard Reserved Instances, but can be exchanged for another Convertible Reserved Instance with different instance attributes. Convertible Reserved Instances can also be modified.

For more information, see [Types of Reserved Instances (Offering Classes)](#).

After you purchase a Reserved Instance, you cannot cancel your purchase. However, you may be able to modify (p. 220), exchange (p. 224), or sell (p. 214) your Reserved Instance if your needs change.
For more information about pricing, see Amazon EC2 Reserved Instances Pricing.

Reserved Instance Limits

There is a limit to the number of Reserved Instances that you can purchase per month. For each Region you can purchase 20 regional (p. 201) Reserved Instances per month plus an additional 20 zonal (p. 201) Reserved Instances per month for each Availability Zone.

For example, in a Region with three Availability Zones, the limit is 80 Reserved Instances per month: 20 regional Reserved Instances for the Region plus 20 zonal Reserved Instances for each of the three Availability Zones (20x3=60).

A regional Reserved Instance applies a discount to a running On-Demand Instance. The default On-Demand Instance limit is 20. You cannot exceed your running On-Demand Instance limit by purchasing regional Reserved Instances. For example, if you already have 20 running On-Demand Instances, and you purchase 20 regional Reserved Instances, the 20 regional Reserved Instances are used to apply a discount to the 20 running On-Demand Instances. If you purchase more regional Reserved Instances, you will not be able to launch more instances because you have reached your On-Demand Instance limit.

Before purchasing regional Reserved Instances, make sure your On-Demand Instance limit matches or exceeds the number of regional Reserved Instances you intend to own. If required, make sure you request an increase to your On-Demand Instance limit before purchasing more regional Reserved Instances.

A zonal Reserved Instance—a Reserved Instance that is purchased for a specific Availability Zone—provides capacity reservation as well as a discount. You can exceed your running On-Demand Instance limit by purchasing zonal Reserved Instances. For example, if you already have 20 running On-Demand Instances, and you purchase 20 zonal Reserved Instances, you can launch a further 20 On-Demand Instances that match the specifications of your zonal Reserved Instances, giving you a total of 40 running instances.

The Amazon EC2 console provides limit information. For more information, see Viewing Your Current Limits (p. 975).

Regional and Zonal Reserved Instances (Scope)

When you purchase a Reserved Instance, you determine the scope of the Reserved Instance. The scope is either regional or zonal.

- **Regional**: When you purchase a Reserved Instance for a Region, it’s referred to as a *regional* Reserved Instance.
- **Zonal**: When you purchase a Reserved Instance for a specific Availability Zone, it’s referred to as a *zonal* Reserved Instance.

Differences between Regional and Zonal Reserved Instances

The following table highlights some key differences between regional Reserved Instances and zonal Reserved Instances:

<table>
<thead>
<tr>
<th></th>
<th>Regional Reserved Instances</th>
<th>Zonal Reserved Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability Zone flexibility</td>
<td>The Reserved Instance discount applies to instance usage in any</td>
<td>No Availability Zone flexibility—the Reserved Instance discount</td>
</tr>
<tr>
<td></td>
<td>Availability Zone in the specified Region.</td>
<td>applies to instance usage in the specified Availability Zone only.</td>
</tr>
</tbody>
</table>

199
Reserved Instances

Regional Reserved Instances | Zonal Reserved Instances
---|---
Capacity reservation | No capacity reservation—a regional Reserved Instance does not provide a capacity reservation. | A zonal Reserved Instance provides a capacity reservation in the specified Availability Zone.

Instance size flexibility | The Reserved Instance discount applies to instance usage within the instance family, regardless of size. Only supported on Amazon Linux/Unix Reserved Instances with default tenancy. For more information, see Instance Size Flexibility Determined by Normalization Factor (p. 201). | No instance size flexibility—the Reserved Instance discount applies to instance usage for the specified instance type and size only.

For more information and examples, see How Reserved Instances Are Applied (p. 200).

Types of Reserved Instances (Offering Classes)

When you purchase a Reserved Instance, you can choose between a Standard or Convertible offering class. The Reserved Instance applies to a single instance type, platform, scope, and tenancy over a term. If your computing needs change, you may be able to modify or exchange your Reserved Instance, depending on the offering class. Offering classes may also have additional restrictions or limitations.

The following are the differences between Standard and Convertible offering classes.

<table>
<thead>
<tr>
<th>Standard Reserved Instance</th>
<th>Convertible Reserved Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some attributes, such as instance size, can be modified during the term; however, the instance family cannot be modified. You cannot exchange a Standard Reserved Instance, only modify it. For more information, see Modifying Reserved Instances (p. 220).</td>
<td>Can be exchanged during the term for another Convertible Reserved Instance with new attributes including instance family, instance type, platform, scope, or tenancy. For more information, see Exchanging Convertible Reserved Instances (p. 224). You can also modify some attributes of a Convertible Reserved Instance. For more information, see Modifying Reserved Instances (p. 220).</td>
</tr>
</tbody>
</table>

Can be sold in the Reserved Instance Marketplace. | Cannot be sold in the Reserved Instance Marketplace. |

Standard and Convertible Reserved Instances can be purchased to apply to instances in a specific Availability Zone (zonal Reserved Instances), or to instances in a Region (regional Reserved Instances). For more information and examples, see How Reserved Instances Are Applied (p. 200).

If you want to purchase capacity reservations that recur on a daily, weekly, or monthly basis, a Scheduled Reserved Instance may meet your needs. For more information, see Scheduled Reserved Instances (p. 227).

How Reserved Instances Are Applied

If you purchase a Reserved Instance and you already have a running instance that matches the specifications of the Reserved Instance, the billing benefit is immediately applied. You do not have to
restart your instances. If you do not have an eligible running instance, launch an instance and ensure that
you match the same criteria that you specified for your Reserved Instance. For more information, see
Using Your Reserved Instances (p. 214).

Reserved Instances apply to usage in the same manner, irrespective of the offering type (Standard or
Convertible), and are automatically applied to running On-Demand Instances with matching attributes.

How Zonal Reserved Instances Are Applied

Reserved Instances assigned to a specific Availability Zone provide the Reserved Instance discount
to matching instance usage in that Availability Zone. For example, if you purchase two c4.xlarge
default tenancy Linux/Unix Standard Reserved Instances in Availability Zone us-east-1a, then up to two
c4.xlarge default tenancy Linux/Unix instances running in the Availability Zone us-east-1a can benefit
from the Reserved Instance discount. The attributes (tenancy, platform, Availability Zone, instance type,
and instance size) of the running instances must match that of the Reserved Instances.

How Regional Reserved Instances Are Applied

Regional Reserved Instances are purchased for a Region and provide Availability Zone flexibility. The
Reserved Instance discount applies to instance usage in any Availability Zone in that Region.

Regional Reserved Instances also provide instance size flexibility where the Reserved Instance discount
applies to instance usage within the instance family, regardless of size.

Limitations for instance size flexibility

Instance size flexibility only applies to Reserved Instances that use the Amazon Linux/Unix platform with
default tenancy.

Instance size flexibility does not apply to the following Reserved Instances:

• Reserved Instances that are purchased for a specific Availability Zone (zonal Reserved Instances)
• Reserved Instances with dedicated tenancy
• Reserved Instances for Windows, Windows with SQL Standard, Windows with SQL Server Enterprise,
  Windows with SQL Server Web, RHEL, and SLES

Instance Size Flexibility Determined by Normalization Factor

Instance size flexibility is determined by the normalization factor of the instance size. The discount
applies either fully or partially to running instances of the same instance family, depending on the
instance size of the reservation, in any Availability Zone in the Region. The only attributes that must be
matched are the instance family, tenancy, and platform.

Instance size flexibility is applied from the smallest to the largest instance size within the instance family
based on the normalization factor.

The following table lists the different sizes within an instance family, and the corresponding
normalization factor per hour. This scale is used to apply the discounted rate of Reserved Instances to
the normalized usage of the instance family.

<table>
<thead>
<tr>
<th>Instance size</th>
<th>Normalization factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>nano</td>
<td>0.25</td>
</tr>
<tr>
<td>micro</td>
<td>0.5</td>
</tr>
</tbody>
</table>
### Instance size vs. Normalization factor

<table>
<thead>
<tr>
<th>Instance size</th>
<th>Normalization factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>1</td>
</tr>
<tr>
<td>medium</td>
<td>2</td>
</tr>
<tr>
<td>large</td>
<td>4</td>
</tr>
<tr>
<td>xlarge</td>
<td>8</td>
</tr>
<tr>
<td>2xlarge</td>
<td>16</td>
</tr>
<tr>
<td>3xlarge</td>
<td>24</td>
</tr>
<tr>
<td>4xlarge</td>
<td>32</td>
</tr>
<tr>
<td>6xlarge</td>
<td>48</td>
</tr>
<tr>
<td>8xlarge</td>
<td>64</td>
</tr>
<tr>
<td>9xlarge</td>
<td>72</td>
</tr>
<tr>
<td>10xlarge</td>
<td>80</td>
</tr>
<tr>
<td>12xlarge</td>
<td>96</td>
</tr>
<tr>
<td>16xlarge</td>
<td>128</td>
</tr>
<tr>
<td>18xlarge</td>
<td>144</td>
</tr>
<tr>
<td>24xlarge</td>
<td>192</td>
</tr>
<tr>
<td>32xlarge</td>
<td>256</td>
</tr>
</tbody>
</table>

For example, a `t2.medium` instance has a normalization factor of 2. If you purchase a `t2.medium` default tenancy Amazon Linux/Unix Reserved Instance in the US East (N. Virginia) and you have two running `t2.small` instances in your account in that Region, the billing benefit is applied in full to both instances.

Or, if you have one `t2.large` instance running in your account in the US East (N. Virginia) Region, the billing benefit is applied to 50% of the usage of the instance.
The normalization factor is also applied when modifying Reserved Instances. For more information, see Modifying Reserved Instances (p. 220).

Normalization Factor for Bare Metal Instances

Instance size flexibility also applies to bare metal instances within the instance family. If you have regional Amazon Linux/Unix Reserved Instances with shared tenancy on bare metal instances, you can benefit from the Reserved Instance savings within the same instance family. The opposite is also true: if you have regional Amazon Linux/Unix Reserved Instances with shared tenancy on instances in the same family as a bare metal instance, you can benefit from the Reserved Instance savings on the bare metal instance.

A bare metal instance is the same size as the largest instance within the same instance family. For example, an i3.metal is the same size as an i3.16xlarge, so they have the same normalization factor.

**Note**
The .metal instance sizes do not have a single normalization factor. They vary based on the specific instance family.

<table>
<thead>
<tr>
<th>Bare metal instance size</th>
<th>Normalization factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>c5.metal</td>
<td>192</td>
</tr>
<tr>
<td>i3.metal</td>
<td>128</td>
</tr>
<tr>
<td>r5.metal</td>
<td>192</td>
</tr>
<tr>
<td>r5d.metal</td>
<td>192</td>
</tr>
<tr>
<td>z1d.metal</td>
<td>96</td>
</tr>
<tr>
<td>m5.metal</td>
<td>192</td>
</tr>
<tr>
<td>m5d.metal</td>
<td>192</td>
</tr>
</tbody>
</table>

For example, an i3.metal instance has a normalization factor of 128. If you purchase an i3.metal default tenancy Amazon Linux/Unix Reserved Instance in the US East (N. Virginia), the billing benefit can apply as follows:

- If you have one running i3.16xlarge in your account in that Region, the billing benefit is applied in full to the i3.16xlarge instance (i3.16xlarge normalization factor = 128).
• Or, if you have two running i3.8xlarge instances in your account in that Region, the billing benefit is applied in full to both i3.8xlarge instances (i3.8xlarge normalization factor = 64).

• Or, if you have four running i3.4xlarge instances in your account in that Region, the billing benefit is applied in full to all four i3.4xlarge instances (i3.4xlarge normalization factor = 32).

The opposite is also true. For example, if you purchase two i3.8xlarge default tenancy Amazon Linux/Unix Reserved Instances in the US East (N. Virginia), and you have one running i3.metal instance in that Region, the billing benefit is applied in full to the i3.metal instance.

Examples of Applying Reserved Instances

The following scenarios cover the ways in which Reserved Instances are applied.

Example Scenario 1: Reserved Instances in a Single Account

You are running the following On-Demand Instances in account A:

• 4 x m3.large Linux, default tenancy instances in Availability Zone us-east-1a
• 2 x m4.xlarge Amazon Linux, default tenancy instances in Availability Zone us-east-1b
• 1 x c4.xlarge Amazon Linux, default tenancy instances in Availability Zone us-east-1c

You purchase the following Reserved Instances in account A:

• 4 x m3.large Linux, default tenancy Reserved Instances in Availability Zone us-east-1a (capacity is reserved)
• 4 x m4.large Amazon Linux, default tenancy Reserved Instances in Region us-east-1
• 1 x c4.large Amazon Linux, default tenancy Reserved Instances in Region us-east-1

The Reserved Instance benefits are applied in the following way:

• The discount and capacity reservation of the four m3.large zonal Reserved Instances is used by the four m3.large instances because the attributes (instance size, Region, platform, tenancy) between them match.

• The m4.large regional Reserved Instances provide Availability Zone and instance size flexibility, because they are regional Amazon Linux Reserved Instances with default tenancy.

An m4.large is equivalent to 4 normalized units/hour.

You've purchased four m4.large regional Reserved Instances, and in total, they are equal to 16 normalized units/hour (4x4). Account A has two m4.xlarge instances running, which is equivalent to 16 normalized units/hour (2x8). In this case, the four m4.large regional Reserved Instances provide the billing benefit to an entire hour of usage of the two m4.xlarge instances.

• The c4.large regional Reserved Instance in us-east-1 provides Availability Zone and instance size flexibility, because it is a regional Amazon Linux Reserved Instance with default tenancy, and applies to the c4.xlarge instance. A c4.large instance is equivalent to 4 normalized units/hour and a c4.xlarge is equivalent to 8 normalized units/hour.

In this case, the c4.large regional Reserved Instance provides partial benefit to c4.xlarge usage. This is because the c4.large Reserved Instance is equivalent to 4 normalized units/hour, but the c4.xlarge instance requires 8 normalized units/hour. Therefore, the c4.large Reserved Instance billing discount applies to 50% of c4.xlarge usage. The remaining c4.xlarge usage is charged at the On-Demand rate.
Example Scenario 2: Regional Reserved Instances in Linked Accounts

Reserved Instances are first applied to usage within the purchasing account, followed by qualifying usage in any other account in the organization. For more information, see Reserved Instances and Consolidated Billing (p. 207). For regional Reserved Instances that offer instance size flexibility, the benefit is applied from the smallest to the largest instance size within the instance family.

You're running the following On-Demand Instances in account A (the purchasing account):

- 2 x m4.xlarge Linux, default tenancy instances in Availability Zone us-east-1a
- 1 x m4.2xlarge Linux, default tenancy instances in Availability Zone us-east-1b
- 2 x c4.xlarge Linux, default tenancy instances in Availability Zone us-east-1a
- 1 x c4.2xlarge Linux, default tenancy instances in Availability Zone us-east-1b

Another customer is running the following On-Demand Instances in account B—a linked account:

- 2 x m4.xlarge Linux, default tenancy instances in Availability Zone us-east-1a

You purchase the following regional Reserved Instances in account A:

- 4 x m4.xlarge Linux, default tenancy Reserved Instances in Region us-east-1
- 2 x c4.xlarge Linux, default tenancy Reserved Instances in Region us-east-1

The regional Reserved Instance benefits are applied in the following way:

- The discount of the four m4.xlarge Reserved Instances is used by the two m4.xlarge instances and the single m4.2xlarge instance in account A (purchasing account). All three instances match the attributes (instance family, Region, platform, tenancy). The discount is applied to instances in the purchasing account (account A) first, even though account B (linked account) has two m4.xlarge that also match the Reserved Instances. There is no capacity reservation because the Reserved Instances are regional Reserved Instances.
- The discount of the two c4.xlarge Reserved Instances applies to the two c4.xlarge instances, because they are a smaller instance size than the c4.2xlarge instance. There is no capacity reservation because the Reserved Instances are regional Reserved Instances.

Example Scenario 3: Zonal Reserved Instances in a Linked Account

In general, Reserved Instances that are owned by an account are applied first to usage in that account. However, if there are qualifying, unused Reserved Instances for a specific Availability Zone (zonal Reserved Instances) in other accounts in the organization, they are applied to the account before regional Reserved Instances owned by the account. This is done to ensure maximum Reserved Instance utilization and a lower bill. For billing purposes, all the accounts in the organization are treated as one account. The following example may help explain this.

You're running the following On-Demand Instance in account A (the purchasing account):

- 1 x m4.xlarge Linux, default tenancy instance in Availability Zone us-east-1a

A customer is running the following On-Demand Instance in linked account B:

- 1 x m4.xlarge Linux, default tenancy instance in Availability Zone us-east-1b

You purchase the following regional Reserved Instances in account A:
• 1 x m4.xlarge Linux, default tenancy Reserved Instance in Region us-east-1

A customer also purchases the following zonal Reserved Instances in linked account C:
• 1 x m4.xlarge Linux, default tenancy Reserved Instances in Availability Zone us-east-1a

The Reserved Instance benefits are applied in the following way:
• The discount of the m4.xlarge zonal Reserved Instance owned by account C is applied to the m4.xlarge usage in account A.
• The discount of the m4.xlarge regional Reserved Instance owned by account A is applied to the m4.xlarge usage in account B.
• If the regional Reserved Instance owned by account A was first applied to the usage in account A, the zonal Reserved Instance owned by account C remains unused and usage in account B is charged at On-Demand rates.

For more information, see Reserved Instances in the Billing and Cost Management Report.

How You Are Billed

All Reserved Instances provide you with a discount compared to On-Demand pricing. With Reserved Instances, you pay for the entire term regardless of actual use. You can choose to pay for your Reserved Instance upfront, partially upfront, or monthly, depending on the payment option (p. 198) specified for the Reserved Instance.

When Reserved Instances expire, you are charged On-Demand rates for EC2 instance usage. You can queue a Reserved Instance for purchase up to three years in advance. This can help you ensure that you have uninterrupted coverage. For more information, see Queuing Your Purchase (p. 210).

The AWS Free Tier is available for new AWS accounts. If you are using the AWS Free Tier to run Amazon EC2 instances, and you purchase a Reserved Instance, you are charged under standard pricing guidelines. For information, see AWS Free Tier.

Contents
• Usage Billing (p. 206)
• Viewing Your Bill (p. 207)
• Reserved Instances and Consolidated Billing (p. 207)
• Reserved Instance Discount Pricing Tiers (p. 207)

Usage Billing

Reserved Instances are billed for every clock-hour during the term that you select, regardless of whether an instance is running. Each clock-hour starts on the hour (zero minutes and zero seconds past the hour) of a standard 24-hour clock. For example, 1:00:00 to 1:59:59 is one clock-hour. For more information about instance states, see Instance Lifecycle (p. 329).

Reserved Instance billing benefits only apply to one instance-hour per clock-hour. An instance-hour begins when an instance is started and continues for 60 minutes or until the instance is stopped or terminated—whichever happens first.

A new instance-hour begins after an instance has run for 60 continuous minutes, or if an instance is stopped and then started. Rebooting an instance does not reset the running instance-hour.

For example, if an instance is stopped and then started again during a clock-hour and continues running for two more clock-hours, the first instance-hour (before the restart) is charged at the discounted
Reserved Instance rate. The next instance-hour (after restart) is charged at the On-Demand rate and the next two instance-hours are charged at the discounted Reserved Instance rate.

Cost Explorer on the Billing and Cost Management console enables you to analyze the savings against running On-Demand Instances. The Reserved Instances FAQ includes an example of a list value calculation.

If you close your AWS account, On-Demand billing for your resources stops. However, if you have any Reserved Instances in your account, you continue to receive a bill for these until they expire.

Viewing Your Bill

You can find out about the charges and fees to your account by viewing the AWS Billing and Cost Management console.

- The Dashboard displays a spend summary for your account.
- On the Bills page, under Details expand the Elastic Compute Cloud section and the Region to get billing information about your Reserved Instances.

You can view the charges online, or you can download a CSV file.

You can also track your Reserved Instance utilization using the AWS Cost and Usage Report. For more information, see Reserved Instances under Cost and Usage Report in the AWS Billing and Cost Management User Guide.

Reserved Instances and Consolidated Billing

The pricing benefits of Reserved Instances are shared when the purchasing account is part of a set of accounts billed under one consolidated billing payer account. The instance usage across all member accounts is aggregated in the payer account every month. This is typically useful for companies in which there are different functional teams or groups; then, the normal Reserved Instance logic is applied to calculate the bill. For more information, see Consolidated Billing and AWS Organizations in the AWS Organizations User Guide.

If you close the payer account, any member accounts that benefit from Reserved Instances billing discounts continue to benefit from the discount until the Reserved Instances expire, or until the member account is removed.

Reserved Instance Discount Pricing Tiers

If your account qualifies for a discount pricing tier, it automatically receives discounts on upfront and instance usage fees for Reserved Instance purchases that you make within that tier level from that point on. To qualify for a discount, the list value of your Reserved Instances in the Region must be $500,000 USD or more.

The following rules apply:

- Pricing tiers and related discounts apply only to purchases of Amazon EC2 Standard Reserved Instances.
- Pricing tiers do not apply to Reserved Instances for Windows with SQL Server Standard, SQL Server Web, and SQL Server Enterprise.
- Pricing tiers do not apply to Reserved Instances for Linux with SQL Server Standard, SQL Server Web, and SQL Server Enterprise.
- Pricing tier discounts only apply to purchases made from AWS. They do not apply to purchases of third-party Reserved Instances.
- Discount pricing tiers are currently not applicable to Convertible Reserved Instance purchases.
Calculating Reserved Instance Pricing Discounts

You can determine the pricing tier for your account by calculating the list value for all of your Reserved Instances in a Region. Multiply the hourly recurring price for each reservation by the total number of hours for the term and add the undiscounted upfront price (also known as the fixed price) listed on the Reserved Instances pricing page at the time of purchase. Because the list value is based on undiscounted (public) pricing, it is not affected if you qualify for a volume discount or if the price drops after you buy your Reserved Instances.

\[
\text{List value} = \text{fixed price} + (\text{undiscounted recurring hourly price} \times \text{hours in term})
\]

For example, for a 1-year Partial Upfront \(\times 2\) \textit{t2.small} Reserved Instance, assume the upfront price is $60.00 and the hourly rate is $0.007. This provides a list value of $121.32.

\[
121.32 = 60.00 + (0.007 \times 8760)
\]

To view the fixed price values for Reserved Instances using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Reserved Instances.
3. Display the Upfront Price column by choosing Show/Hide Columns (the gear-shaped icon) in the top right corner.

To view the fixed price values for Reserved Instances using the command line

- \texttt{describe-reserved-instances} (AWS CLI)
- \texttt{Get-EC2ReservedInstance} (AWS Tools for Windows PowerShell)
- \texttt{DescribeReservedInstances} (Amazon EC2 API)

Buying with a Discount Tier

When you buy Reserved Instances, Amazon EC2 automatically applies any discounts to the part of your purchase that falls within a discount pricing tier. You don't need to do anything differently, and you can buy Reserved Instances using any of the Amazon EC2 tools. For more information, see Buying Reserved Instances (p. 209).

After the list value of your active Reserved Instances in a Region crosses into a discount pricing tier, any future purchase of Reserved Instances in that Region are charged at a discounted rate. If a single purchase of Reserved Instances in a Region takes you over the threshold of a discount tier, then the portion of the purchase that is above the price threshold is charged at the discounted rate. For more information about the temporary Reserved Instance IDs that are created during the purchase process, see Crossing Pricing Tiers (p. 209).

If your list value falls below the price point for that discount pricing tier—for example, if some of your Reserved Instances expire—future purchases of Reserved Instances in the Region are not discounted. However, you continue to get the discount applied against any Reserved Instances that were originally purchased within the discount pricing tier.
When you buy Reserved Instances, one of four possible scenarios occurs:

- **No discount**—Your purchase within a Region is still below the discount threshold.
- **Partial discount**—Your purchase within a Region crosses the threshold of the first discount tier. No discount is applied to one or more reservations and the discounted rate is applied to the remaining reservations.
- **Full discount**—Your entire purchase within a Region falls within one discount tier and is discounted appropriately.
- **Two discount rates**—Your purchase within a Region crosses from a lower discount tier to a higher discount tier. You are charged two different rates: one or more reservations at the lower discounted rate, and the remaining reservations at the higher discounted rate.

**Crossing Pricing Tiers**

If your purchase crosses into a discounted pricing tier, you see multiple entries for that purchase: one for that part of the purchase charged at the regular price, and another for that part of the purchase charged at the applicable discounted rate.

The Reserved Instance service generates several Reserved Instance IDs because your purchase crossed from an undiscounted tier, or from one discounted tier to another. There is an ID for each set of reservations in a tier. Consequently, the ID returned by your purchase CLI command or API action is different from the actual ID of the new Reserved Instances.

**Consolidated Billing for Pricing Tiers**

A consolidated billing account aggregates the list value of member accounts within a Region. When the list value of all active Reserved Instances for the consolidated billing account reaches a discount pricing tier, any Reserved Instances purchased after this point by any member of the consolidated billing account are charged at the discounted rate (as long as the list value for that consolidated account stays above the discount pricing tier threshold). For more information, see Reserved Instances and Consolidated Billing (p. 207).

**Buying Reserved Instances**

To purchase a Reserved Instance, search for *Reserved Instance offerings* from AWS and third-party sellers, adjusting your search parameters until you find the exact match that you're looking for.

When you search for Reserved Instances to buy, you receive a quote on the cost of the returned offerings. When you proceed with the purchase, AWS automatically places a limit price on the purchase price. The total cost of your Reserved Instances won't exceed the amount that you were quoted.

If the price rises or changes for any reason, the purchase is not completed. If, at the time of purchase, there are offerings similar to your choice but at a lower price, AWS sells you the offerings at the lower price.

Before you confirm your purchase, review the details of the Reserved Instance that you plan to buy, and make sure that all the parameters are accurate. After you purchase a Reserved Instance (either from a third-party seller in the Reserved Instance Marketplace or from AWS), you cannot cancel your purchase.

**Note**

To purchase and modify Reserved Instances, ensure that your IAM user account has the appropriate permissions, such as the ability to describe Availability Zones. For information, see Example Policies for Working With the AWS CLI or an AWS SDK and Example Policies for Working in the Amazon EC2 Console.

**Tasks**

- **Choosing a Platform** (p. 210)
Choosing a Platform

When you purchase a Reserved Instance, you must choose an offering for a platform that represents the operating system for your instance.

For Windows with SQL Standard, Windows with SQL Server Enterprise, and Windows with SQL Server Web, you must choose offerings for those specific platforms. For all other Windows versions, choose an offering for the Windows platform.

Queuing Your Purchase

By default, when you purchase a Reserved Instance, it is executed immediately. Alternatively, you can queue your purchases for a future date and time. For example, you can queue a purchase for around the time that an existing Reserved Instance expires. This can help you ensure that you have uninterrupted coverage.

You can queue purchases for regional Reserved Instances, but not zonal Reserved Instances or Reserved Instances from other sellers. You can queue a purchase up to three years in advance. On the scheduled date and time, the purchase is executed using the default payment method. After the payment is successful, the billing benefit is applied.

You can view your queued purchases in the Amazon EC2 console. The status of a queued purchase is queued. You can cancel a queued purchase any time before its scheduled time. For details, see Canceling a Queued Purchase (p. 213).

Buying Standard Reserved Instances

You can buy Standard Reserved Instances in a specific Availability Zone and get a capacity reservation. Alternatively, you can forego the capacity reservation and purchase a regional Standard Reserved Instance.

To buy Standard Reserved Instances using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Reserved Instances, Purchase Reserved Instances.
3. For Offering Class, choose Standard to display Standard Reserved Instances.
4. To purchase a capacity reservation, choose Only show offerings that reserve capacity in the top-right corner of the purchase screen. To purchase a regional Reserved Instance, leave the check box unselected.
5. Select other configurations as needed and choose Search.

To purchase a Standard Reserved Instance from the Reserved Instance Marketplace, look for 3rd Party in the Seller column in the search results. The Term column displays non-standard terms.
6. Select the Reserved Instances to purchase, enter the quantity, and choose Add to Cart.
7. To see a summary of the Reserved Instances that you selected, choose View Cart.
8. If Order On is Now, the purchase is completed immediately. To queue a purchase, select a future date and time. You can select a different date for each eligible offering in the cart.
9. To complete the order, choose **Order**.

If, at the time of placing the order, there are offerings similar to your choice but with a lower price, AWS sells you the offerings at the lower price.

10. The status of your order is listed in the **State** column. When your order is complete, the **State** value changes from payment–pending to active. When the Reserved Instance is active, it is ready to use.

**Note**

If the status goes to retired, AWS may not have received your payment.

**To buy a Standard Reserved Instance using the AWS CLI**

1. Find available Reserved Instances using the `describe-reserved-instances-offerings` command. Specify `standard` for the --offering-class parameter to return only Standard Reserved Instances. You can apply additional parameters to narrow your results; for example, if you want to purchase a regional `t2.large` Reserved Instance with a default tenancy for Linux/UNIX for a 1-year term only:

   ```
   aws ec2 describe-reserved-instances-offerings --instance-type t2.large --offering-class standard --product-description "Linux/UNIX" --instance-tenancy default --filters Name=duration,Values=31536000 Name=scope,Values=Region
   ```

To find Reserved Instances on the Reserved Instance Marketplace only, use the `marketplace` filter and do not specify a duration in the request, as the term may be shorter than a 1– or 3-year term.

   ```
   aws ec2 describe-reserved-instances-offerings --instance-type t2.large --offering-class standard --product-description "Linux/UNIX" --instance-tenancy default --filters Name=marketplace,Values=true
   ```

When you find a Reserved Instance that meets your needs, take note of the offering ID. For example:

   ```
   "ReservedInstancesOfferingId": "bec624df-a8cc-4aad-a72f-4f8abc34caf2"
   ```

2. Use the `purchase-reserved-instances-offering` command to buy your Reserved Instance. You must specify the Reserved Instance offering ID you obtained the previous step and you must specify the number of instances for the reservation.

   ```
   aws ec2 purchase-reserved-instances-offering --reserved-instances-offering-id bec624df-a8cc-4aad-a72f-4f8abc34caf2 --instance-count 1
   ```

3. Use the `describe-reserved-instances` command to get the status of your Reserved Instance.

   ```
   aws ec2 describe-reserved-instances
   ```

Alternatively, use the following AWS Tools for Windows PowerShell commands:

- `Get-EC2ReservedInstancesOffering`
- `New-EC2ReservedInstance`
- `Get-EC2ReservedInstance`

After the purchase is complete, if you already have a running instance that matches the specifications of the Reserved Instance, the billing benefit is immediately applied. You do not have to restart your
instances. If you do not have a suitable running instance, launch an instance and ensure that you match
the same criteria that you specified for your Reserved Instance. For more information, see Using Your
Reserved Instances (p. 214).

For examples of how Reserved Instances are applied to your running instances, see How Reserved
Instances Are Applied (p. 200).

Buying Convertible Reserved Instances

You can buy Convertible Reserved Instances in a specific Availability Zone and get a capacity reservation.
Alternatively, you can forego the capacity reservation and purchase a regional Convertible Reserved
Instance.

To buy Convertible Reserved Instances using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Reserved Instances, Purchase Reserved Instances.
3. For Offering Class, choose Convertible to display Convertible Reserved Instances.
4. To purchase a capacity reservation, choose Only show offerings that reserve capacity in the top-
right corner of the purchase screen. To purchase a regional Reserved Instance, leave the check box
unselected.
5. Select other configurations as needed and choose Search.
6. Select the Convertible Reserved Instances to purchase, enter the quantity, and choose Add to Cart.
7. To see a summary of your selection, choose View Cart.
8. If Order On is Now, the purchase is completed immediately. To queue a purchase, select a future
date and time. You can select a different date for each eligible offering in the cart.
9. To complete the order, choose Order.

   If, at the time of placing the order, there are offerings similar to your choice but with a lower price,
AWS sells you the offerings at the lower price.
10. The status of your order is listed in the State column. When your order is complete, the State value
changes from payment-pending to active. When the Reserved Instance is active, it is ready to
use.

Note
If the status goes to retired, AWS may not have received your payment.

To buy a Convertible Reserved Instance using the AWS CLI

1. Find available Reserved Instances using the describe-reserved-instances-offerings command.
Specify convertible for the --offering-class parameter to return only Convertible Reserved
Instances. You can apply additional parameters to narrow your results; for example, if you want to
purchase a regional t2.large Reserved Instance with a default tenancy for Linux/UNIX:

   aws ec2 describe-reserved-instances-offerings --instance-type t2.large --offering-
   class convertible --product-description "Linux/UNIX" --instance-tenancy default --
   filters Name=scope,Values=Region

   When you find a Reserved Instance that meets your needs, take note of the offering ID. For example:

   "ReservedInstancesOfferingId": "bec624df-a8cc-4aad-a72f-4f8abc34caf2"

2. Use the purchase-reserved-instances-offering command to buy your Reserved Instance. You must
specify the Reserved Instance offering ID you obtained the previous step and you must specify the
number of instances for the reservation.
3. Use the `describe-reserved-instances` command to get the status of your Reserved Instance.

```
aws ec2 describe-reserved-instances
```

Alternatively, use the following AWS Tools for Windows PowerShell commands:

- `Get-EC2ReservedInstancesOffering`
- `New-EC2ReservedInstance`
- `Get-EC2ReservedInstance`

If you already have a running instance that matches the specifications of the Reserved Instance, the billing benefit is immediately applied. You do not have to restart your instances. If you do not have a suitable running instance, launch an instance and ensure that you match the same criteria that you specified for your Reserved Instance. For more information, see Using Your Reserved Instances (p. 214).

For examples of how Reserved Instances are applied to your running instances, see How Reserved Instances Are Applied (p. 200).

### Viewing Your Reserved Instances

You can view the Reserved Instances you've purchased using the Amazon EC2 console, or a command line tool.

**To view your Reserved Instances in the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Reserved Instances**.
3. Your active and retired Reserved Instances are listed. The **State** column displays the state.
4. If you are a seller in the Reserved Instance Marketplace the **My Listings** tab displays the status of a reservation that's listed in the Reserved Instance Marketplace (p. 214). For more information, see Reserved Instance Listing States (p. 218).

**To view your Reserved Instances using the command line**

- `describe-reserved-instances` (AWS CLI)
- `Get-EC2ReservedInstance` (Tools for Windows PowerShell)

### Canceling a Queued Purchase

You can queue a purchase up to three years in advance. You can cancel a queued purchase any time before its scheduled time.

**To cancel a queued purchase**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Reserved Instances**.
3. Select one or more Reserved Instances.
4. Choose **Actions, Delete Queued Reserved Instances**.
5. When prompted for confirmation, choose **Yes, Delete**.
Using Your Reserved Instances

Reserved Instances are automatically applied to running On-Demand Instances provided that the specifications match. If you have no running On-Demand Instances that match the specifications of your Reserved Instance, the Reserved Instance is unused until you launch an instance with the required specifications.

If you're launching an instance to take advantage of the billing benefit of a Reserved Instance, ensure that you specify the following information during launch:

- Platform: You must choose an Amazon Machine Image (AMI) that matches the platform (product description) of your Reserved Instance. For example, if you specified Linux/UNIX, you can launch an instance from an Amazon Linux AMI or an Ubuntu AMI.
- Instance type: Specify the same instance type as your Reserved Instance; for example, t2.large.
- Availability Zone: If you purchased a Reserved Instance for a specific Availability Zone, you must launch the instance into the same Availability Zone. If you purchased a regional Reserved Instance, you can launch your instance into any Availability Zone.
- Tenancy: The tenancy of your instance must match the tenancy of the Reserved Instance; for example, dedicated or shared. For more information, see Dedicated Instances (p. 312).

For more information, see Launching an Instance Using the Launch Instance Wizard (p. 334). For examples of how Reserved Instances are applied to your running instances, see How Reserved Instances Are Applied (p. 200).

You can use Amazon EC2 Auto Scaling or other AWS services to launch the On-Demand Instances that use your Reserved Instance benefits. For more information, see the Amazon EC2 Auto Scaling User Guide.

Reserved Instance Marketplace

The Reserved Instance Marketplace is a platform that supports the sale of third-party and AWS customers' unused Standard Reserved Instances, which vary in term lengths and pricing options.

For example, you may want to sell Reserved Instances after moving instances to a new AWS Region, changing to a new instance type, ending projects before the term expiration, when your business needs change, or if you have unneeded capacity.

If you want to sell your unused Reserved Instances on the Reserved Instance Marketplace, you must meet certain eligibility criteria.

Contents
- Selling on the Reserved Instance Marketplace (p. 214)
- Buying from the Reserved Instance Marketplace (p. 220)

Selling on the Reserved Instance Marketplace

As soon as you list your Reserved Instances in the Reserved Instance Marketplace, they are available for potential buyers to find. All Reserved Instances are grouped according to the duration of the term remaining and the hourly price.

To fulfill a buyer's request, AWS first sells the Reserved Instance with the lowest upfront price in the specified grouping. Then, we sell the Reserved Instance with the next lowest price, until the buyer's entire order is fulfilled. AWS then processes the transactions and transfers ownership of the Reserved Instances to the buyer.

You own your Reserved Instance until it's sold. After the sale, you've given up the capacity reservation and the discounted recurring fees. If you continue to use your instance, AWS charges you the On-Demand price starting from the time that your Reserved Instance was sold.
Restrictions and Limitations

Before you can sell your unused reservations, you must register as a seller in the Reserved Instance Marketplace. For information, see Registering as a Seller (p. 215).

The following limitations and restrictions apply when selling Reserved Instances:

- Only Amazon EC2 Standard Reserved Instances can be sold in the Reserved Instance Marketplace. Convertible Reserved Instances cannot be sold. There must be at least one month remaining in the term of the Standard Reserved Instance.
- The minimum price allowed in the Reserved Instance Marketplace is $0.00.
- You can sell No Upfront, Partial Upfront, or All Upfront Reserved Instances in the Reserved Instance Marketplace. If there is an upfront payment on a Reserved Instance, it can be sold only after AWS has received the upfront payment and the reservation has been active (you've owned it) for at least 30 days.
- You cannot modify your listing in the Reserved Instance Marketplace directly. However, you can change your listing by first canceling it and then creating another listing with new parameters. For information, see Pricing Your Reserved Instances (p. 217). You can also modify your Reserved Instances before listing them. For information, see Modifying Reserved Instances (p. 220).
- AWS charges a service fee of 12 percent of the total upfront price of each Standard Reserved Instance you sell in the Reserved Instance Marketplace. The upfront price is the price the seller is charging for the Standard Reserved Instance.
- Only Amazon EC2 Standard Reserved Instances can be sold in the Reserved Instance Marketplace. Other AWS Reserved Instances, such as Amazon RDS and Amazon ElastiCache Reserved Instances cannot be sold in the Reserved Instance Marketplace.

Registering as a Seller

Note

Only the AWS account root user can register an account as a seller.

To sell in the Reserved Instance Marketplace, you must first register as a seller. During registration, you provide the following information:

- **Bank information**—AWS must have your bank information in order to disburse funds collected when you sell your reservations. The bank you specify must have a US address. For more information, see Bank Account for Disbursement (p. 216).
- **Tax information**—All sellers are required to complete a tax information interview to determine any necessary tax reporting obligations. For more information, see Tax Information (p. 216).
After AWS receives your completed seller registration, you receive an email confirming your registration and informing you that you can get started selling in the Reserved Instance Marketplace.

**Bank Account for Disbursement**

AWS must have your bank information in order to disburse funds collected when you sell your Reserved Instance. The bank you specify must have a US address.

**To register a default bank account for disbursements**

1. Open the [Reserved Instance Marketplace Seller Registration](#) page and sign in using your AWS credentials.
2. On the **Manage Bank Account** page, provide the following information about the bank through to receive payment:
   - Bank account holder name
   - Routing number
   - Account number
   - Bank account type

   **Note**
   If you are using a corporate bank account, you are prompted to send the information about the bank account via fax (1-206-765-3424).

After registration, the bank account provided is set as the default, pending verification with the bank. It can take up to two weeks to verify a new bank account, during which time you can’t receive disbursements. For an established account, it usually takes about two days for disbursements to complete.

**To change the default bank account for disbursement**

1. On the **Reserved Instance Marketplace Seller Registration** page, sign in with the account that you used when you registered.
2. On the **Manage Bank Account** page, add a new bank account or modify the default bank account as needed.

**Tax Information**

Your sale of Reserved Instances might be subject to a transaction-based tax, such as sales tax or value-added tax. You should check with your business's tax, legal, finance, or accounting department to determine if transaction-based taxes are applicable. You are responsible for collecting and sending the transaction-based taxes to the appropriate tax authority.

As part of the seller registration process, you must complete a tax interview in the **Seller Registration Portal**. The interview collects your tax information and populates an IRS form W-9, W-8BEN, or W-8BEN-E, which is used to determine any necessary tax reporting obligations.

The tax information you enter as part of the tax interview might differ depending on whether you operate as an individual or business, and whether you or your business are a US or non-US person or entity. As you fill out the tax interview, keep in mind the following:

- Information provided by AWS, including the information in this topic, does not constitute tax, legal, or other professional advice. To find out how the IRS reporting requirements might affect your business, or if you have other questions, contact your tax, legal, or other professional advisor.
- To fulfill the IRS reporting requirements as efficiently as possible, answer all questions and enter all information requested during the interview.
Check your answers. Avoid misspellings or entering incorrect tax identification numbers. They can result in an invalidated tax form.

Based on your tax interview responses and IRS reporting thresholds, Amazon may file Form 1099-K. Amazon mails a copy of your Form 1099-K on or before January 31 in the year following the year that your tax account reaches the threshold levels. For example, if your account reaches the threshold in 2018, your Form 1099-K is mailed on or before January 31, 2019.

For more information about IRS requirements and Form 1099-K, see the IRS website.

Pricing Your Reserved Instances

The upfront fee is the only fee that you can specify for the Reserved Instance that you're selling. The upfront fee is the one-time fee that the buyer pays when they purchase a Reserved Instance. You cannot specify the usage fee or the recurring fee; The buyer pays the same usage or recurring fees that were set when the reservations were originally purchased.

The following are important limits to note:

- **You can sell up to $50,000 in Reserved Instances per year.** To sell more, complete the Request to Raise Sales Limit on Amazon EC2 Reserved Instances form.
- **The minimum price is $0.** The minimum allowed price in the Reserved Instance Marketplace is $0.00.

You cannot modify your listing directly. However, you can change your listing by first canceling it and then creating another listing with new parameters.

You can cancel your listing at any time, as long as it's in the active state. You cannot cancel the listing if it's already matched or being processed for a sale. If some of the instances in your listing are matched and you cancel the listing, only the remaining unmatched instances are removed from the listing.

Because the value of Reserved Instances decreases over time, by default, AWS can set prices to decrease in equal increments month over month. However, you can set different upfront prices based on when your reservation sells.

For example, if your Reserved Instance has nine months of its term remaining, you can specify the amount that you would accept if a customer were to purchase that Reserved Instance with nine months remaining. You could set another price with five months remaining, and yet another price with one month remaining.

Listing Your Reserved Instances

As a registered seller, you can choose to sell one or more of your Reserved Instances. You can choose to sell all of them in one listing or in portions. In addition, you can list Reserved Instances with any configuration of instance type, platform, and scope.

The console determines a suggested price. It checks for offerings that match your Reserved Instance and matches the one with the lowest price. Otherwise, it calculates a suggested price based on the cost of the Reserved Instance for its remaining time. If the calculated value is less than $1.01, the suggested price is $1.01.

If you cancel your listing and a portion of that listing has already been sold, the cancellation is not effective on the portion that has been sold. Only the unsold portion of the listing is no longer available in the Reserved Instance Marketplace.

To list a Reserved Instance in the Reserved Instance Marketplace using the AWS Management Console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Reserved Instances**.
3. Select the Reserved Instances to list, and choose **Sell Reserved Instances**.
4. On the **Configure Your Reserved Instance Listing** page, set the number of instances to sell and the upfront price for the remaining term in the relevant columns. See how the value of your reservation changes over the remainder of the term by selecting the arrow next to the **Months Remaining** column.
5. If you are an advanced user and you want to customize the pricing, you can enter different values for the subsequent months. To return to the default linear price drop, choose **Reset**.
6. Choose **Continue** when you are finished configuring your listing.
7. Confirm the details of your listing, on the **Confirm Your Reserved Instance Listing** page and if you're satisfied, choose **List Reserved Instance**.

**To view your listings in the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Reserved Instances**.
3. Select the Reserved Instance that you've listed and choose **My Listings**.

**To manage Reserved Instances in the Reserved Instance Marketplace using the AWS CLI**

1. Get a list of your Reserved Instances by using the **describe-reserved-instances** command.
2. Note the ID of the Reserved Instance you want to list and call **create-reserved-instances-listing**. You must specify the ID of the Reserved Instance, the number of instances, and the pricing schedule.
3. To view your listing, use the **describe-reserved-instances-listings** command.
4. To cancel your listing, use the **cancel-reserved-instances-listings** command.

**Reserved Instance Listing States**

**Listing State** on the **My Listings** tab of the Reserved Instances page displays the current status of your listings:

The information displayed by **Listing State** is about the status of your listing in the Reserved Instance Marketplace. It is different from the status information that is displayed by the **State** column in the **Reserved Instances** page. This **State** information is about your reservation.

- **active**—The listing is available for purchase.
- **canceled**—The listing is canceled and isn't available for purchase in the Reserved Instance Marketplace.
- **closed**—The Reserved Instance is not listed. A Reserved Instance might be closed because the sale of the listing was completed.

**Lifecycle of a Listing**

When all the instances in your listing are matched and sold, the **My Listings** tab shows that the **Total instance count** matches the count listed under **Sold**. Also, there are no **Available** instances left for your listing, and its **Status** is **closed**.

When only a portion of your listing is sold, AWS retires the Reserved Instances in the listing and creates the number of Reserved Instances equal to the Reserved Instances remaining in the count. So, the listing ID and the listing that it represents, which now has fewer reservations for sale, is still active.

Any future sales of Reserved Instances in this listing are processed this way. When all the Reserved Instances in the listing are sold, AWS marks the listing as **closed**.
For example, you create a listing `Reserved Instances listing ID Sec28771-05ff-4b9b-aa31-9e57dex ample` with a listing count of 5.

The **My Listings** tab in the **Reserved Instance** console page displays the listing this way:

`Reserved Instance listing ID Sec28771-05ff-4b9b-aa31-9e57dex ample`

- Total reservation count = 5
- Sold = 0
- Available = 5
- Status = active

A buyer purchases two of the reservations, which leaves a count of three reservations still available for sale. Because of this partial sale, AWS creates a new reservation with a count of three to represent the remaining reservations that are still for sale.

This is how your listing looks in the **My Listings** tab:

`Reserved Instance listing ID Sec28771-05ff-4b9b-aa31-9e57dex ample`

- Total reservation count = 5
- Sold = 2
- Available = 3
- Status = active

If you cancel your listing and a portion of that listing has already sold, the cancelation is not effective on the portion that has been sold. Only the unsold portion of the listing is no longer available in the Reserved Instance Marketplace.

**After Your Reserved Instance Is Sold**

When your Reserved Instance is sold, AWS sends you an email notification. Each day that there is any kind of activity, you receive one email notification capturing all the activities of the day. For example, you create or sell a listing, or AWS sends funds to your account.

To track the status of a Reserved Instance listing in the console, choose **Reserved Instance, My Listings**. The **My Listings** tab contains the **Listing State** value. It also contains information about the term, listing price, and a breakdown of how many instances in the listing are available, pending, sold, and canceled. You can also use the `describe-reserved-instances-listings` command with the appropriate filter to obtain information about your listings.

**Getting Paid**

As soon as AWS receives funds from the buyer, a message is sent to the registered owner account email for the sold Reserved Instance.

AWS sends an Automated Clearing House (ACH) wire transfer to your specified bank account. Typically, this transfer occurs between one to three days after your Reserved Instance has been sold. Disbursements take place once a day. You will receive an email with a disbursement report after the funds are released. Keep in mind that you can't receive disbursements until AWS receives verification from your bank. This can take up to two weeks.

The Reserved Instance that you sold continues to appear when you describe your Reserved Instances.

You receive a cash disbursement for your Reserved Instances through a wire transfer directly into your bank account. AWS charges a service fee of 12 percent of the total upfront price of each Reserved Instance you sell in the Reserved Instance Marketplace.
Information Shared with the Buyer

When you sell in the Reserved Instance Marketplace, AWS shares your company's legal name on the buyer's statement in accordance with US regulations. In addition, if the buyer calls AWS Support because the buyer needs to contact you for an invoice or for some other tax-related reason, AWS may need to provide the buyer with your email address so that the buyer can contact you directly.

For similar reasons, the buyer’s ZIP code and country information are provided to the seller in the disbursement report. As a seller, you might need this information to accompany any necessary transaction taxes that you remit to the government (such as sales tax and value-added tax).

AWS cannot offer tax advice, but if your tax specialist determines that you need specific additional information, contact AWS Support.

Buying from the Reserved Instance Marketplace

You can purchase Reserved Instances from third-party sellers who own Reserved Instances that they no longer need from the Reserved Instance Marketplace. You can do this using the Amazon EC2 console or a command line tool. The process is similar to purchasing Reserved Instances from AWS. For more information, see Buying Reserved Instances (p. 209).

There are a few differences between Reserved Instances purchased in the Reserved Instance Marketplace and Reserved Instances purchased directly from AWS:

- **Term**—Reserved Instances that you purchase from third-party sellers have less than a full standard term remaining. Full standard terms from AWS run for one year or three years.
- **Upfront price**—Third-party Reserved Instances can be sold at different upfront prices. The usage or recurring fees remain the same as the fees set when the Reserved Instances were originally purchased from AWS.
- **Types of Reserved Instances**—Only Amazon EC2 Standard Reserved Instances can be purchased from the Reserved Instance Marketplace. Convertible Reserved Instances, Amazon RDS and Amazon ElastiCache Reserved Instances are not available for purchase on the Reserved Instance Marketplace.

Basic information about you is shared with the seller, for example, your ZIP code and country information.

This information enables sellers to calculate any necessary transaction taxes that they have to remit to the government (such as sales tax or value-added tax) and is provided as a disbursement report. In rare circumstances, AWS might have to provide the seller with your email address, so that they can contact you regarding questions related to the sale (for example, tax questions).

For similar reasons, AWS shares the legal entity name of the seller on the buyer's purchase invoice. If you need additional information about the seller for tax or related reasons, contact AWS Support.

Modifying Reserved Instances

When your needs change, you can modify your Standard or Convertible Reserved Instances and continue to benefit from the billing benefit. You can modify attributes such as the Availability Zone and scope of your Reserved Instance.

**Note**

You can also exchange a Convertible Reserved Instance for another Convertible Reserved Instance with a different configuration. For more information, see Exchanging Convertible Reserved Instances (p. 224).

After modification, the benefit of the Reserved Instances is applied only to instances that match the new parameters. For example, if you change the Availability Zone of a reservation, the capacity reservation
and pricing benefits are automatically applied to instance usage in the new Availability Zone. Instances that no longer match the new parameters are charged at the On-Demand rate, unless your account has other applicable reservations.

If your modification request succeeds:

- The modified reservation becomes effective immediately and the pricing benefit is applied to the new instances beginning at the hour of the modification request. For example, if you successfully modify your reservations at 9:15PM, the pricing benefit transfers to your new instance at 9:00PM. You can get the effective date of the modified Reserved Instances by using the describe-reserved-instances command.
- The original reservation is retired. Its end date is the start date of the new reservation, and the end date of the new reservation is the same as the end date of the original Reserved Instance. If you modify a three-year reservation that had 16 months left in its term, the resulting modified reservation is a 16-month reservation with the same end date as the original one.
- The modified reservation lists a $0 fixed price and not the fixed price of the original reservation.
- The fixed price of the modified reservation does not affect the discount pricing tier calculations applied to your account, which are based on the fixed price of the original reservation.

If your modification request fails, your Reserved Instances maintain their original configuration, and are immediately available for another modification request.

There is no fee for modification, and you do not receive any new bills or invoices.

You can modify your reservations as frequently as you like, but you cannot change or cancel a pending modification request after you submit it. After the modification has completed successfully, you can submit another modification request to roll back any changes you made, if needed.

Contents

- Requirements and Restrictions for Modification (p. 221)
- Submitting Modification Requests (p. 222)
- Troubleshooting Modification Requests (p. 223)

Requirements and Restrictions for Modification

You can modify these attributes as follows.

<table>
<thead>
<tr>
<th>Modifiable attribute</th>
<th>Supported platforms</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change Availability Zones within the same Region</td>
<td>Linux and Windows</td>
<td>-</td>
</tr>
<tr>
<td>Change the scope from Availability Zone to Region and vice versa</td>
<td>Linux and Windows</td>
<td>If you change the scope from Availability Zone to Region, you lose the capacity reservation benefit. If you change the scope from Region to Availability Zone, you lose Availability Zone flexibility and instance size flexibility (if applicable). For more information, see How Reserved Instances Are Applied (p. 200).</td>
</tr>
</tbody>
</table>
Reserved Instances

<table>
<thead>
<tr>
<th>Modifiable attribute</th>
<th>Supported platforms</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change the instance size within the same instance family</td>
<td>Linux only</td>
<td>The reservation must use default tenancy. Some instance families are not supported, because there are no other sizes available. For more information, see Support for Modifying Instance Sizes in the Amazon EC2 User Guide for Linux Instances.</td>
</tr>
<tr>
<td>Change the network from EC2-Classic to Amazon VPC and vice versa</td>
<td>Linux and Windows</td>
<td>The network platform must be available in your AWS account. If you created your AWS account after 2013-12-04, it does not support EC2-Classic.</td>
</tr>
</tbody>
</table>

Requirements

Amazon EC2 processes your modification request if there is sufficient capacity for your target configuration (if applicable), and if the following conditions are met:

- The Reserved Instance cannot be modified before or at the same time that you purchase it
- The Reserved Instance must be active
- There cannot be a pending modification request
- The Reserved Instance is not listed in the Reserved Instance Marketplace
- The input Reserved Instances are all Standard Reserved Instances or all Convertible Reserved Instances, not some of each type
- The input Reserved Instances must expire within the same hour, if they are Standard Reserved Instances

Submitting Modification Requests

Before you modify your Reserved Instances, ensure that you have read the applicable restrictions (p. 221).

To modify your Reserved Instances using the AWS Management Console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the Reserved Instances page, select one or more Reserved Instances to modify, and choose Modify Reserved Instances.
   
   **Note**
   
   If your Reserved Instances are not in the active state or cannot be modified, Modify Reserved Instances is disabled.

3. The first entry in the modification table displays attributes of selected Reserved Instances, and at least one target configuration beneath it. The Units column displays the total instance size footprint. Choose Add for each new configuration to add. Modify the attributes as needed for each configuration, and then choose Continue:
   
   - **Scope**: Choose whether the Reserved Instance applies to an Availability Zone or to the whole Region.
   - **Availability Zone**: Choose the required Availability Zone. Not applicable for regional Reserved Instances.
• **Count**: Specify the number of instances to be covered by the reservation.

4. To confirm your modification choices when you finish specifying your target configurations, choose **Submit Modifications**. If you change your mind at any point, choose **Cancel** to exit the wizard.

5. You can determine the status of your modification request by looking at the **State** column in the Reserved Instances screen. The following are the possible states.

- **active (pending modification)** — Transition state for original Reserved Instances
- **retired (pending modification)** — Transition state for original Reserved Instances while new Reserved Instances are being created
- **retired** — Reserved Instances successfully modified and replaced
- **active** — One of the following:
  - New Reserved Instances created from a successful modification request
  - Original Reserved Instances after a failed modification request

**To modify your Reserved Instances using the command line**

1. To modify your Reserved Instances, you can use one of the following commands:
   - `modify-reserved-instances` (AWS CLI)
   - `Edit-EC2ReservedInstance` (AWS Tools for Windows PowerShell)

2. To get the status of your modification request (processing, fulfilled, or failed), use one of the following commands:
   - `describe-reserved-instances-modifications` (AWS CLI)
   - `Get-EC2ReservedInstancesModification` (AWS Tools for Windows PowerShell)

**Troubleshooting Modification Requests**

If the target configuration settings that you requested were unique, you receive a message that your request is being processed. At this point, Amazon EC2 has only determined that the parameters of your modification request are valid. Your modification request can still fail during processing due to unavailable capacity.

In some situations, you might get a message indicating incomplete or failed modification requests instead of a confirmation. Use the information in such messages as a starting point for resubmitting another modification request. Ensure that you have read the applicable restrictions (p. 221) before submitting the request.

**Not all selected Reserved Instances can be processed for modification**

Amazon EC2 identifies and lists the Reserved Instances that cannot be modified. If you receive a message like this, go to the **Reserved Instances** page in the Amazon EC2 console and check the information for the Reserved Instances.

**Error in processing your modification request**

You submitted one or more Reserved Instances for modification and none of your requests can be processed. Depending on the number of reservations you are modifying, you can get different versions of the message.

Amazon EC2 displays the reasons why your request cannot be processed. For example, you might have specified the same target configuration—a combination of Availability Zone and platform—for one or more subsets of the Reserved Instances you are modifying. Try submitting the modification requests again, but ensure that the instance details of the reservations match, and that the target configurations for all subsets being modified are unique.
Exchanging Convertible Reserved Instances

You can exchange one or more Convertible Reserved Instances for another Convertible Reserved Instance with a different configuration, including instance family, operating system, and tenancy. There are no limits to how many times you perform an exchange, as long as the target Convertible Reserved Instance is of an equal or higher value than the Convertible Reserved Instances that you are exchanging.

When you exchange your Convertible Reserved Instance, the number of instances for your current reservation is exchanged for a number of instances that cover the equal or higher value of the configuration of the target Convertible Reserved Instance. Amazon EC2 calculates the number of Reserved Instances that you can receive as a result of the exchange.

Contents

- Requirements for Exchanging Convertible Reserved Instances (p. 224)
- Calculating Convertible Reserved Instances Exchanges (p. 225)
- Merging Convertible Reserved Instances (p. 225)
- Exchanging a Portion of a Convertible Reserved Instance (p. 226)
- Submitting Exchange Requests (p. 226)

Requirements for Exchanging Convertible Reserved Instances

If the following conditions are met, Amazon EC2 processes your exchange request. Your Convertible Reserved Instance must be:

- Active
- Not pending a previous exchange request

The following rules apply:

- Convertible Reserved Instances can only be exchanged for other Convertible Reserved Instances currently offered by AWS.
- Convertible Reserved Instances are associated with a specific Region, which is fixed for the duration of the reservation's term. You cannot exchange a Convertible Reserved Instance for a Convertible Reserved Instance in a different Region.
- You can exchange one or more Convertible Reserved Instances at a time for one Convertible Reserved Instance only.
- To exchange a portion of a Convertible Reserved Instance, you can modify it into two or more reservations, and then exchange one or more of the reservations for a new Convertible Reserved Instance. For more information, see Exchanging a Portion of a Convertible Reserved Instance (p. 226).
  For more information about modifying your Reserved Instances, see Modifying Reserved Instances (p. 220).
- All Upfront Convertible Reserved Instances can be exchanged for Partial Upfront Convertible Reserved Instances, and vice versa.

  **Note**
  If the total upfront payment required for the exchange (true-up cost) is less than $0.00, AWS automatically gives you a quantity of instances in the Convertible Reserved Instance that ensures that true-up cost is $0.00 or more.

  **Note**
  If the total value (upfront price + hourly price * number of remaining hours) of the new Convertible Reserved Instance is less than the total value of the exchanged Convertible Reserved Instance, AWS automatically gives you a quantity of instances in the Convertible Reserved Instance that ensures that the total value is the same or higher than that of the exchanged Convertible Reserved Instance.
To benefit from better pricing, you can exchange a No Upfront Convertible Reserved Instance for an All Upfront or Partial Upfront Convertible Reserved Instance.

You cannot exchange All Upfront and Partial Upfront Convertible Reserved Instances for No Upfront Convertible Reserved Instances.

You can exchange a No Upfront Convertible Reserved Instance for another No Upfront Convertible Reserved Instance only if the new Convertible Reserved Instance's hourly price is the same or higher than the exchanged Convertible Reserved Instance's hourly price.

**Note**

If the total value (hourly price * number of remaining hours) of the new Convertible Reserved Instance is less than the total value of the exchanged Convertible Reserved Instance, AWS automatically gives you a quantity of instances in the Convertible Reserved Instance that ensures that the total value is the same or higher than that of the exchanged Convertible Reserved Instance.

- If you exchange multiple Convertible Reserved Instances that have different expiration dates, the expiration date for the new Convertible Reserved Instance is the date that's furthest in the future.
- If you exchange a single Convertible Reserved Instance, it must have the same term (1-year or 3-years) as the new Convertible Reserved Instance. If you merge multiple Convertible Reserved Instances with different term lengths, the new Convertible Reserved Instance has a 3-year term. For more information, see Merging Convertible Reserved Instances (p. 225).

**Calculating Convertible Reserved Instances Exchanges**

Exchanging Convertible Reserved Instances is free. However, you may be required to pay a true-up cost, which is a prorated upfront cost of the difference between the Convertible Reserved Instances that you had and the Convertible Reserved Instances that you receive from the exchange.

Each Convertible Reserved Instance has a list value. This list value is compared to the list value of the Convertible Reserved Instances that you want in order to determine how many instance reservations you can receive from the exchange.

For example: You have 1 x $35-list value Convertible Reserved Instance that you want to exchange for a new instance type with a list value of $10.

\[
\frac{35}{10} = 3.5
\]

You can exchange your Convertible Reserved Instance for three $10 Convertible Reserved Instances. It's not possible to purchase half reservations; therefore you must purchase an additional Convertible Reserved Instance to cover the remainder:

\[3.5 = 3 \text{ whole Convertible Reserved Instances} + 1 \text{ additional Convertible Reserved Instance.}\]

The fourth Convertible Reserved Instance has the same end date as the other three. If you are exchanging Partial or All Upfront Convertible Reserved Instances, you pay the true-up cost for the fourth reservation. If the remaining upfront cost of your Convertible Reserved Instances is $500, and the target reservation would normally cost $600 on a prorated basis, you are charged $100.

\[
\frac{600 \text{ prorated upfront cost of new reservations} - 500 \text{ remaining upfront cost of original reservations}}{100 \text{ difference}} = \frac{100}{100} = 1
\]

**Merging Convertible Reserved Instances**

If you merge two or more Convertible Reserved Instances, the term of the new Convertible Reserved Instance must be the same as the original Convertible Reserved Instances, or the highest of the original Convertible Reserved Instances. The expiration date for the new Convertible Reserved Instance is the expiration date that's furthest in the future.
For example, you have the following Convertible Reserved Instances in your account:

<table>
<thead>
<tr>
<th>Reserved Instance ID</th>
<th>Term</th>
<th>Expiration date</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaaa1111</td>
<td>1-year</td>
<td>2018-12-31</td>
</tr>
<tr>
<td>bbbb2222</td>
<td>1-year</td>
<td>2018-07-31</td>
</tr>
<tr>
<td>cccc3333</td>
<td>3-year</td>
<td>2018-06-30</td>
</tr>
<tr>
<td>dddd4444</td>
<td>3-year</td>
<td>2019-12-31</td>
</tr>
</tbody>
</table>

- You can merge `aaaa1111` and `bbbb2222` and exchange them for a 1-year Convertible Reserved Instance. You cannot exchange them for a 3-year Convertible Reserved Instance. The expiration date of the new Convertible Reserved Instance is 2018-12-31.
- You can merge `bbbb2222` and `cccc3333` and exchange them for a 3-year Convertible Reserved Instance. You cannot exchange them for a 1-year Convertible Reserved Instance. The expiration date of the new Convertible Reserved Instance is 2018-07-31.
- You can merge `cccc3333` and `dddd4444` and exchange them for a 3-year Convertible Reserved Instance. You cannot exchange them for a 1-year Convertible Reserved Instance. The expiration date of the new Convertible Reserved Instance is 2019-12-31.

### Exchanging a Portion of a Convertible Reserved Instance

You can use the modification process to split your Convertible Reserved Instance into smaller reservations, and then exchange one or more of the new reservations for a new Convertible Reserved Instance. The following examples demonstrate how you can do this.

**Example Example: Convertible Reserved Instance with multiple instances**

In this example, you have a `t2.micro` Convertible Reserved Instance with four instances in the reservation. To exchange two `t2.micro` instances for an `m4.xlarge` instance:

1. Modify the `t2.micro` Convertible Reserved Instance by splitting it into two `t2.micro` Convertible Reserved Instances with two instances each.
2. Exchange one of the new `t2.micro` Convertible Reserved Instances for an `m4.xlarge` Convertible Reserved Instance.

### Submitting Exchange Requests

You can exchange your Convertible Reserved Instances using the Amazon EC2 console or a command line tool.
Exchanging a Convertible Reserved Instance Using the Console

You can search for Convertible Reserved Instances offerings and select your new configuration from the choices provided.

To exchange Convertible Reserved Instances using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Reserved Instances, select the Convertible Reserved Instances to exchange, and choose Actions, Exchange Reserved Instance.
3. Select the attributes of the desired configuration using the drop-down menus, and choose Find Offering.
4. Select a new Convertible Reserved Instance. The Instance Count column displays the number of Reserved Instances that you receive for the exchange. When you have selected a Convertible Reserved Instance that meets your needs, choose Exchange.

The Reserved Instances that were exchanged are retired, and the new Reserved Instances are displayed in the Amazon EC2 console. This process can take a few minutes to propagate.

Exchanging a Convertible Reserved Instance Using the Command Line Interface

To exchange a Convertible Reserved Instance, first find a target Convertible Reserved Instance that meets your needs:

• describe-reserved-instances-offerings (AWS CLI)
• Get-EC2ReservedInstancesOffering (Tools for Windows PowerShell)

Get a quote for the exchange, which includes the number of Reserved Instances you get from the exchange, and the true-up cost for the exchange:

• get-reserved-instances-exchange-quote (AWS CLI)
• GetEC2-ReservedInstancesExchangeQuote (Tools for Windows PowerShell)

Finally, perform the exchange:

• accept-reserved-instances-exchange-quote (AWS CLI)
• Confirm-EC2ReservedInstancesExchangeQuote (Tools for Windows PowerShell)

Scheduled Reserved Instances

Scheduled Reserved Instances (Scheduled Instances) enable you to purchase capacity reservations that recur on a daily, weekly, or monthly basis, with a specified start time and duration, for a one-year term. You reserve the capacity in advance, so that you know it is available when you need it. You pay for the time that the instances are scheduled, even if you do not use them.

Scheduled Instances are a good choice for workloads that do not run continuously, but do run on a regular schedule. For example, you can use Scheduled Instances for an application that runs during business hours or for batch processing that runs at the end of the week.

If you require a capacity reservation on a continuous basis, Reserved Instances might meet your needs and decrease costs. For more information, see Reserved Instances (p. 197). If you are flexible about when your instances run, Spot Instances might meet your needs and decrease costs. For more information, see Spot Instances (p. 231).
How Scheduled Instances Work

Amazon EC2 sets aside pools of EC2 instances in each Availability Zone for use as Scheduled Instances. Each pool supports a specific combination of instance type, operating system, and network.

To get started, you must search for an available schedule. You can search across multiple pools or a single pool. After you locate a suitable schedule, purchase it.

You must launch your Scheduled Instances during their scheduled time periods, using a launch configuration that matches the following attributes of the schedule that you purchased: instance type, Availability Zone, network, and platform. When you do so, Amazon EC2 launches EC2 instances on your behalf, based on the specified launch specification. Amazon EC2 must ensure that the EC2 instances have terminated by the end of the current scheduled time period so that the capacity is available for any other Scheduled Instances it is reserved for. Therefore, Amazon EC2 terminates the EC2 instances three minutes before the end of the current scheduled time period.

You can't stop or reboot Scheduled Instances, but you can terminate them manually as needed. If you terminate a Scheduled Instance before its current scheduled time period ends, you can launch it again after a few minutes. Otherwise, you must wait until the next scheduled time period.

The following diagram illustrates the lifecycle of a Scheduled Instance.

Service-Linked Roles for Scheduled Instances

Amazon EC2 creates a service-linked role when you purchase a Scheduled Instance. A service-linked role includes all the permissions that Amazon EC2 requires to call other AWS services on your behalf. For more information, see Using Service-Linked Roles in the IAM User Guide.

Amazon EC2 uses the service-linked role named AWSServiceRoleForEC2ScheduledInstances to complete the following actions:

- `ec2:TerminateInstances` - Terminate Scheduled Instances after their schedules complete
- `ec2:CreateTags` - Add system tags to Scheduled Instances
If you purchased Scheduled Instances before October 2017, when Amazon EC2 began supporting this service-linked role, Amazon EC2 created the `AWSServiceRoleForEC2ScheduledInstances` role in your AWS account. For more information, see A New Role Appeared in My Account in the IAM User Guide.

If you no longer need to use Scheduled Instances, we recommend that you delete the `AWSServiceRoleForEC2ScheduledInstances` role. After this role is deleted from your account, Amazon EC2 will create the role again if you purchase Scheduled Instances.

### Purchasing a Scheduled Instance

To purchase a Scheduled Instance, you can use the Scheduled Reserved Instances Reservation Wizard.

**Warning**

After you purchase a Scheduled Instance, you can't cancel, modify, or resell your purchase.

#### To purchase a Scheduled Instance (console)

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, under **INSTANCES**, choose **Scheduled Instances**. If the currently selected Region does not support Scheduled Instances, the page is unavailable. Learn more (p. 230)
3. Choose **Purchase Scheduled Instances**.
4. On the **Find available schedules** page, do the following:
   a. Under **Create a schedule**, select the starting date from **Starting on**, the schedule recurrence (daily, weekly, or monthly) from **Recurring**, and the minimum duration from **for duration**. Note that the console ensures that you specify a value for the minimum duration that meets the minimum required utilization for your Scheduled Instance (1,200 hours per year).

   ![Create a schedule](image)

   a. Under **Instance details**, select the operating system and network from **Platform**. To narrow the results, select one or more instance types from **Instance type** or one or more Availability Zones from **Availability Zone**.

   ![Instance details](image)

   b. Under **Available schedules**, select one or more schedules. For each schedule that you select, set the quantity of instances and choose **Add to Cart**.

   c. Under **Review and purchase** page, verify your selections and edit them as needed. When you are finished, choose **Purchase**.

#### To purchase a Scheduled Instance (AWS CLI)

Use the `describe-scheduled-instance-availability` command to list the available schedules that meet your needs, and then use the `purchase-scheduled-instances` command to complete the purchase.
Launching a Scheduled Instance

After you purchase a Scheduled Instance, it is available for you to launch during its scheduled time periods.

To launch a Scheduled Instance (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, under **INSTANCES**, choose **Scheduled Instances**. If the currently selected Region does not support Scheduled Instances, the page is unavailable. Learn more (p. 230)
3. Select the Scheduled Instance and choose **Launch Scheduled Instances**.
4. On the **Configure** page, complete the launch specification for your Scheduled Instances and choose **Review**.
   
   **Important**
   
   The launch specification must match the instance type, Availability Zone, network, and platform of the schedule that you purchased.
5. On the **Review** page, verify the launch configuration and modify it as needed. When you are finished, choose **Launch**.

To launch a Scheduled Instance (AWS CLI)

Use the `describe-scheduled-instances` command to list your Scheduled Instances, and then use the `run-scheduled-instances` command to launch each Scheduled Instance during its scheduled time periods.

Scheduled Instance Limits

Scheduled Instances are subject to the following limits:

- The following are the only supported instance types: C3, C4, M4, and R3.
- The required term is 365 days (one year).
- The minimum required utilization is 1,200 hours per year.
- You can purchase a Scheduled Instance up to three months in advance.
- They are available in the following Regions: US East (N. Virginia), US West (Oregon), and Europe (Ireland).
Spot Instances

A Spot Instance is an unused EC2 instance that is available for less than the On-Demand price. Because Spot Instances enable you to request unused EC2 instances at steep discounts, you can lower your Amazon EC2 costs significantly. The hourly price for a Spot Instance is called a Spot price. The Spot price of each instance type in each Availability Zone is set by Amazon EC2, and adjusted gradually based on the long-term supply of and demand for Spot Instances. Your Spot Instance runs whenever capacity is available and the maximum price per hour for your request exceeds the Spot price.

Spot Instances are a cost-effective choice if you can be flexible about when your applications run and if your applications can be interrupted. For example, Spot Instances are well-suited for data analysis, batch jobs, background processing, and optional tasks. For more information, see Amazon EC2 Spot Instances.

Topics

- Concepts (p. 231)
- How to Get Started (p. 232)
- Related Services (p. 233)
- Pricing and Savings (p. 233)

Concepts

Before you get started with Spot Instances, you should be familiar with the following concepts:

- **Spot Instance pool** – A set of unused EC2 instances with the same instance type (for example, m5.large), operating system, Availability Zone, and network platform.
- **Spot price** – The current price of a Spot Instance per hour.
- **Spot Instance request** – Provides the maximum price per hour that you are willing to pay for a Spot Instance. If you don't specify a maximum price, the default maximum price is the On-Demand price. When the maximum price per hour for your request exceeds the Spot price, Amazon EC2 fulfills your request if capacity is available. A Spot Instance request is either **one-time** or **persistent**. Amazon EC2 automatically resubmits a persistent Spot request after the Spot Instance associated with the request is terminated. Your Spot Instance request can optionally specify a duration for the Spot Instances.
- **Spot Fleet** – A set of Spot Instances that is launched based on criteria that you specify. The Spot Fleet selects the Spot Instance pools that meet your needs and launches Spot Instances to meet the target capacity for the fleet. By default, Spot Fleets are set to maintain target capacity by launching replacement instances after Spot Instances in the fleet are terminated. You can submit a Spot Fleet as a one-time request, which does not persist after the instances have been terminated. You can include On-Demand Instance requests in a Spot Fleet request.
- **Spot Instance interruption** – Amazon EC2 terminates, stops, or hibernates your Spot Instance when the Spot price exceeds the maximum price for your request or capacity is no longer available. Amazon EC2 provides a Spot Instance interruption notice, which gives the instance a two-minute warning before it is interrupted.

Key Differences between Spot Instances and On-Demand Instances

The following table lists the key differences between Spot Instances and On-Demand Instances.

<table>
<thead>
<tr>
<th></th>
<th>Spot Instances</th>
<th>On-Demand Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Launch time</td>
<td>Can only be launched immediately if the Spot Request is active and capacity is available.</td>
<td>Can only be launched immediately if you make a manual launch request and capacity is available.</td>
</tr>
</tbody>
</table>
Strategies for Using Spot Instances

One strategy to maintain a minimum level of guaranteed compute resources for your applications is to launch a core group of On-Demand Instances, and supplement them with Spot Instances when the opportunity arises.

Another strategy is to launch Spot Instances with a specified duration (also known as Spot blocks), which are designed not to be interrupted and will run continuously for the duration you select. In rare situations, Spot blocks may be interrupted due to Amazon EC2 capacity needs. In these cases, we provide a two-minute warning before we terminate an instance, and you are not charged for the terminated instances even if you used them. For more information, see Specifying a Duration for Your Spot Instances (p. 246).

How to Get Started

The first thing you need to do is get set up to use Amazon EC2. It can also be helpful to have experience launching On-Demand Instances before launching Spot Instances.

Get up and running

- Setting Up with Amazon EC2 (p. 12)
- Getting Started with Amazon EC2 Windows Instances (p. 20)

Spot basics

- How Spot Instances Work (p. 234)
- How Spot Fleet Works (p. 235)
Working with Spot Instances

- Preparing for Interruptions (p. 288)
- Creating a Spot Instance Request (p. 248)
- Getting Request Status Information (p. 283)

Working with Spot Fleets

- Spot Fleet Prerequisites (p. 255)
- Creating a Spot Fleet Request (p. 259)

Related Services

You can provision Spot Instances directly using Amazon EC2. You can also provision Spot Instances using other services in AWS. For more information, see the following documentation.

Amazon EC2 Auto Scaling and Spot Instances

You can create launch configurations with the maximum price that you are willing to pay, so that Amazon EC2 Auto Scaling can launch Spot Instances. For more information, see Launching Spot Instances in Your Auto Scaling Group and Using Multiple Instance Types and Purchase Options in the Amazon EC2 Auto Scaling User Guide.

Amazon EMR and Spot Instances

There are scenarios where it can be useful to run Spot Instances in an Amazon EMR cluster. For more information, see Spot Instances and When Should You Use Spot Instances in the Amazon EMR Management Guide.

AWS CloudFormation Templates

AWS CloudFormation enables you to create and manage a collection of AWS resources using a template in JSON format. AWS CloudFormation templates can include the maximum price you are willing to pay. For more information, see EC2 Spot Instance Updates - Auto Scaling and CloudFormation Integration.

AWS SDK for Java

You can use the Java programming language to manage your Spot Instances. For more information, see Tutorial: Amazon EC2 Spot Instances and Tutorial: Advanced Amazon EC2 Spot Request Management.

AWS SDK for .NET

You can use the .NET programming environment to manage your Spot Instances. For more information, see Tutorial: Amazon EC2 Spot Instances.

Pricing and Savings

You pay the Spot price for Spot Instances, which is set by Amazon EC2 and adjusted gradually based on the long-term supply of and demand for Spot Instances. If the maximum price for your request exceeds the current Spot price, Amazon EC2 fulfills your request if capacity is available. Your Spot Instances run until you terminate them, capacity is no longer available, the Spot price exceeds your maximum price, or your Amazon EC2 Auto Scaling group terminates them during scale in.

Spot Instances with a predefined duration use a fixed hourly price that remains in effect for the Spot Instance while it runs.
If you or Amazon EC2 interrupts a running Spot Instance, you are charged for the seconds used or the full hour, or you receive no charge, depending on the operating system used and who interrupted the Spot Instance. For more information, see Billing for Interrupted Spot Instances (p. 290).

**View Prices**

To view the current (updated every five minutes) lowest Spot price per region and instance type, see the Spot Instances Pricing page.

To view the Spot price history for the past three months, use the Amazon EC2 console or the describe-spot-price-history command (AWS CLI). For more information, see Spot Instance Pricing History (p. 242).

We independently map Availability Zones to codes for each AWS account. Therefore, you can get different results for the same Availability Zone code (for example, us-west-2a) between different accounts.

**View Savings**

You can view the savings made from using Spot Instances for a single Spot Fleet or for all Spot Instances. You can view the savings made in the last hour or the last three days, and you can view the average cost per vCPU hour and per memory (GiB) hour. Savings are estimated and may differ from actual savings because they do not include the billing adjustments for your usage. For more information about viewing savings information, see Savings From Purchasing Spot Instances (p. 243).

**View Billing**

To review your bill, go to your AWS Account Activity page. Your bill contains links to usage reports that provide details about your bill. For more information, see AWS Account Billing.

If you have questions concerning AWS billing, accounts, and events, contact AWS Support.

**How Spot Instances Work**

To use Spot Instances, create a Spot Instance request or a Spot Fleet request. The request can include the maximum price that you are willing to pay per hour per instance (the default is the On-Demand price), and other constraints such as the instance type and Availability Zone. If your maximum price exceeds the current Spot price for the specified instance, and capacity is available, your request is fulfilled immediately. Otherwise, the request is fulfilled whenever the maximum price exceeds the Spot price and the capacity is available. Spot Instances run until you terminate them or until Amazon EC2 must interrupt them (known as a Spot Instance interruption).

When you use Spot Instances, you must be prepared for interruptions. Amazon EC2 can interrupt your Spot Instance when the Spot price exceeds your maximum price, when the demand for Spot Instances rises, or when the supply of Spot Instances decreases. When Amazon EC2 interrupts a Spot Instance, it provides a Spot Instance interruption notice, which gives the instance a two-minute warning before Amazon EC2 interrupts it. You can't enable termination protection for Spot Instances. For more information, see Spot Instance Interruptions (p. 285).

You can't stop and start an Amazon EBS-backed instance if it is a Spot Instance (only the Spot service can stop and start a Spot Instance), but you can reboot or terminate a Spot Instance.

**Contents**

- Launching Spot Instances in a Launch Group (p. 235)
- Launching Spot Instances in an Availability Zone Group (p. 235)
- Launching Spot Instances in a VPC (p. 235)
Launching Spot Instances in a Launch Group

Specify a launch group in your Spot Instance request to tell Amazon EC2 to launch a set of Spot Instances only if it can launch them all. In addition, if the Spot service must terminate one of the instances in a launch group (for example, if the Spot price exceeds your maximum price), it must terminate them all. However, if you terminate one or more of the instances in a launch group, Amazon EC2 does not terminate the remaining instances in the launch group.

Although this option can be useful, adding this constraint can decrease the chances that your Spot Instance request is fulfilled and increase the chances that your Spot Instances are terminated. For example, your launch group includes instances in multiple Availability Zones. If capacity in one of these Availability Zones decreases and is no longer available, then Amazon EC2 terminates all instances for the launch group.

If you create another successful Spot Instance request that specifies the same (existing) launch group as an earlier successful request, then the new instances are added to the launch group. Subsequently, if an instance in this launch group is terminated, all instances in the launch group are terminated, which includes instances launched by the first and second requests.

Launching Spot Instances in an Availability Zone Group

Specify an Availability Zone group in your Spot Instance request to tell the Spot service to launch a set of Spot Instances in the same Availability Zone. Amazon EC2 need not interrupt all instances in an Availability Zone group at the same time. If Amazon EC2 must interrupt one of the instances in an Availability Zone group, the others remain running.

Although this option can be useful, adding this constraint can lower the chances that your Spot Instance request is fulfilled.

If you specify an Availability Zone group but don’t specify an Availability Zone in the Spot Instance request, the result depends on the network you specified.

Default VPC

Amazon EC2 uses the Availability Zone for the specified subnet. If you don’t specify a subnet, it selects an Availability Zone and its default subnet, but not necessarily the lowest-priced zone. If you deleted the default subnet for an Availability Zone, then you must specify a different subnet.

Nondefault VPC

Amazon EC2 uses the Availability Zone for the specified subnet.

Launching Spot Instances in a VPC

You specify a subnet for your Spot Instances the same way that you specify a subnet for your On-Demand Instances.

- You should use the default maximum price (the On-Demand price), or base your maximum price on the Spot price history of Spot Instances in a VPC.
- [Default VPC] If you want your Spot Instance launched in a specific low-priced Availability Zone, you must specify the corresponding subnet in your Spot Instance request. If you do not specify a subnet, Amazon EC2 selects one for you, and the Availability Zone for this subnet might not have the lowest Spot price.
- [Nondefault VPC] You must specify the subnet for your Spot Instance.

How Spot Fleet Works

A Spot Fleet is a collection, or fleet, of Spot Instances, and optionally On-Demand Instances.
The Spot Fleet attempts to launch the number of Spot Instances and On-Demand Instances to meet the target capacity that you specified in the Spot Fleet request. The request for Spot Instances is fulfilled if there is available capacity and the maximum price you specified in the request exceeds the current Spot price. The Spot Fleet also attempts to maintain its target capacity fleet if your Spot Instances are interrupted.

You can also set a maximum amount per hour that you're willing to pay for your fleet, and Spot Fleet launches instances until it reaches the maximum amount. When the maximum amount you're willing to pay is reached, the fleet stops launching instances even if it hasn't met the target capacity.

A Spot Instance pool is a set of unused EC2 instances with the same instance type (for example, m5.large), operating system, Availability Zone, and network platform. When you make a Spot Fleet request, you can include multiple launch specifications, that vary by instance type, AMI, Availability Zone, or subnet. The Spot Fleet selects the Spot Instance pools that are used to fulfill the request, based on the launch specifications included in your Spot Fleet request, and the configuration of the Spot Fleet request. The Spot Instances come from the selected pools.

Contents
  • On-Demand in Spot Fleet (p. 236)
  • Allocation Strategy for Spot Instances (p. 236)
  • Spot Price Overrides (p. 238)
  • Control Spending (p. 238)
  • Spot Fleet Instance Weighting (p. 239)
  • Walkthrough: Using Spot Fleet with Instance Weighting (p. 240)

On-Demand in Spot Fleet

To ensure that you always have instance capacity, you can include a request for On-Demand capacity in your Spot Fleet request. In your Spot Fleet request, you specify your desired target capacity and how much of that capacity must be On-Demand. The balance comprises Spot capacity, which is launched if there is available Amazon EC2 capacity and availability. For example, if in your Spot Fleet request you specify target capacity as 10 and On-Demand capacity as 8, Amazon EC2 launches 8 capacity units as On-Demand, and 2 capacity units (10-8=2) as Spot.

Prioritizing Instance Types for On-Demand Capacity

When Spot Fleet attempts to fulfill your On-Demand capacity, it defaults to launching the lowest-priced instance type first. If OnDemandAllocationStrategy is set to prioritized, Spot Fleet uses priority to determine which instance type to use first in fulfilling On-Demand capacity. The priority is assigned to the launch template override, and the highest priority is launched first.

For example, you have configured three launch template overrides, each with a different instance type: c3.large, c4.large, and c5.large. The On-Demand price for c5.large is less than for c4.large. c3.large is the cheapest. If you do not use priority to determine the order, the fleet fulfills On-Demand capacity by starting with c3.large, and then c5.large. Because you often have unused Reserved Instances for c4.large, you can set the launch template override priority so that the order is c4.large, c3.large, and then c5.large.

Allocation Strategy for Spot Instances

The allocation strategy for the Spot Instances in your Spot Fleet determines how it fulfills your Spot Fleet request from the possible Spot Instance pools represented by its launch specifications. The following are the allocation strategies that you can specify in your Spot Fleet request:

lowestPrice

  The Spot Instances come from the pool with the lowest price. This is the default strategy.
The Spot Instances are distributed across all pools.

The Spot Instances come from the pool with optimal capacity for the number of instances that are launching.

The Spot Instances are distributed across the number of Spot pools that you specify. This parameter is valid only when used in combination with `lowestPrice`.

**Maintaining Target Capacity**

After Spot Instances are terminated due to a change in the Spot price or available capacity of a Spot Instance pool, a Spot Fleet of type `maintain` launches replacement Spot Instances. If the allocation strategy is `lowestPrice`, the fleet launches replacement instances in the pool where the Spot price is currently the lowest. If the allocation strategy is `diversified`, the fleet distributes the replacement Spot Instances across the remaining pools. If the allocation strategy is `lowestPrice` in combination with `InstancePoolsToUseCount`, the fleet selects the Spot pools with the lowest price and launches Spot Instances across the number of Spot pools that you specify.

**Configuring Spot Fleet for Cost Optimization**

To optimize the costs for your use of Spot Instances, specify the `lowestPrice` allocation strategy so that Spot Fleet automatically deploys the least expensive combination of instance types and Availability Zones based on the current Spot price.

For On-Demand Instance target capacity, Spot Fleet always selects the least expensive instance type based on the public On-Demand price, while continuing to follow the allocation strategy (either `lowestPrice`, `capacityOptimized`, or `diversified`) for Spot Instances.

**Configuring Spot Fleet for Cost Optimization and Diversification**

To create a fleet of Spot Instances that is both cheap and diversified, use the `lowestPrice` allocation strategy in combination with `InstancePoolsToUseCount`. Spot Fleet automatically deploys the cheapest combination of instance types and Availability Zones based on the current Spot price across the number of Spot pools that you specify. This combination can be used to avoid the most expensive Spot Instances.

**Configuring Spot Fleet for Capacity Optimization**

With Spot Instances, pricing changes slowly over time based on long-term trends in supply and demand, but capacity fluctuates in real time. The `capacityOptimized` strategy automatically launches Spot Instances into the most available pools by looking at real-time capacity data and predicting which are the most available. This works well for workloads such as big data and analytics, image and media rendering, machine learning, and high performance computing that may have a higher cost of interruption associated with restarting work and checkpointing. By offering the possibility of fewer interruptions, the `capacityOptimized` strategy can lower the overall cost of your workload.

**Choosing an Appropriate Allocation Strategy**

You can optimize your Spot Fleets based on your use case.

If your fleet is small or runs for a short time, the probability that your Spot Instances may be interrupted is low, even with all the instances in a single Spot Instance pool. Therefore, the `lowestPrice` strategy is likely to meet your needs while providing the lowest cost.
If your fleet is large or runs for a long time, you can improve the availability of your fleet by distributing
the Spot Instances across multiple pools. For example, if your Spot Fleet request specifies 10 pools and
a target capacity of 100 instances, the fleet launches 10 Spot Instances in each pool. If the Spot price
for one pool exceeds your maximum price for this pool, only 10% of your fleet is affected. Using this
strategy also makes your fleet less sensitive to increases in the Spot price in any one pool over time.

With the diversified strategy, the Spot Fleet does not launch Spot Instances into any pools with a
Spot price that is equal to or higher than the On-Demand price.

To create a cheap and diversified fleet, use the lowestPrice strategy in combination with
InstancePoolsToUseCount. You can use a low or high number of Spot pools across which to allocate
your Spot Instances. For example, if you run batch processing, we recommend specifying a low number
of Spot pools (for example, InstancePoolsToUseCount=2) to ensure that your queue always has
compute capacity while maximizing savings. If you run a web service, we recommend specifying a high
number of Spot pools (for example, InstancePoolsToUseCount=10) to minimize the impact if a Spot
Instance pool becomes temporarily unavailable.

If your fleet runs workloads that may have a higher cost of interruption associated with restarting work
and checkpointing, then use the capacityOptimized strategy. This strategy offers the possibility of
fewer interruptions, which can lower the overall cost of your workload.

**Spot Price Overrides**

Each Spot Fleet request can include a global maximum price, or use the default (the On-Demand price).
Spot Fleet uses this as the default maximum price for each of its launch specifications.

You can optionally specify a maximum price in one or more launch specifications. This price is specific
to the launch specification. If a launch specification includes a specific price, the Spot Fleet uses this
maximum price, overriding the global maximum price. Any other launch specifications that do not
include a specific maximum price still use the global maximum price.

**Control Spending**

Spot Fleet stops launching instances when it has either reached the target capacity or the maximum
amount you're willing to pay. To control the amount you pay per hour for your fleet, you can specify the
SpotMaxTotalPrice for Spot Instances and the OnDemandMaxTotalPrice for On-Demand Instances.
When the maximum total price is reached, Spot Fleet stops launching instances even if it hasn't met the
target capacity.

The following examples show two different scenarios. In the first, Spot Fleet stops launching instances
when it has met the target capacity. In the second, Spot Fleet stops launching instances when it has
reached the maximum amount you're willing to pay.

**Example: Stop launching instances when target capacity is reached**

Given a request for m4.large On-Demand Instances, where:

- On-Demand Price: $0.10 per hour
- OnDemandTargetCapacity: 10
- OnDemandMaxTotalPrice: $1.50

Spot Fleet launches 10 On-Demand Instances because the total of $1.00 (10 instances x $0.10) does not
exceed the OnDemandMaxTotalPrice of $1.50.

**Example: Stop launching instances when maximum total price is reached**

Given a request for m4.large On-Demand Instances, where:
• On-Demand Price: $0.10 per hour
• OnDemandTargetCapacity: 10
• OnDemandMaxTotalPrice: $0.80

If Spot Fleet launches the On-Demand target capacity (10 On-Demand Instances), the total cost per hour would be $1.00. This is more than the amount ($0.80) specified for OnDemandMaxTotalPrice. To prevent spending more than you’re willing to pay, Spot Fleet launches only 8 On-Demand Instances (below the On-Demand target capacity) because launching more would exceed the OnDemandMaxTotalPrice.

**Spot Fleet Instance Weighting**

When you request a fleet of Spot Instances, you can define the capacity units that each instance type would contribute to your application’s performance, and adjust your maximum price for each Spot Instance pool accordingly using *instance weighting*.

By default, the price that you specify is *per instance hour*. When you use the instance weighting feature, the price that you specify is *per unit hour*. You can calculate your price per unit hour by dividing your price for an instance type by the number of units that it represents. Spot Fleet calculates the number of Spot Instances to launch by dividing the target capacity by the instance weight. If the result isn’t an integer, the Spot Fleet rounds it up to the next integer, so that the size of your fleet is not below its target capacity. Spot Fleet can select any pool that you specify in your launch specification, even if the capacity of the instances launched exceeds the requested target capacity.

The following tables provide examples of calculations to determine the price per unit for a Spot Fleet request with a target capacity of 10.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Instance weight</th>
<th>Price per instance hour</th>
<th>Price per unit hour</th>
<th>Number of instances launched</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.xlarge</td>
<td>2</td>
<td>$0.05</td>
<td>.025</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.05 divided by 2)</td>
<td></td>
<td>(10 divided by 2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Instance weight</th>
<th>Price per instance hour</th>
<th>Price per unit hour</th>
<th>Number of instances launched</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.8xlarge</td>
<td>8</td>
<td>$0.10</td>
<td>.0125</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.10 divided by 8)</td>
<td></td>
<td>(10 divided by 8, result rounded up)</td>
</tr>
</tbody>
</table>

Use Spot Fleet instance weighting as follows to provision the target capacity that you want in the pools with the lowest price per unit at the time of fulfillment:

1. Set the target capacity for your Spot Fleet either in instances (the default) or in the units of your choice, such as virtual CPUs, memory, storage, or throughput.
2. Set the price per unit.
3. For each launch configuration, specify the weight, which is the number of units that the instance type represents toward the target capacity.

**Instance Weighting Example**
Consider a Spot Fleet request with the following configuration:

- A target capacity of 24
- A launch specification with an instance type \texttt{r3.2xlarge} and a weight of 6
- A launch specification with an instance type \texttt{c3.xlarge} and a weight of 5

The weights represent the number of units that instance type represents toward the target capacity. If the first launch specification provides the lowest price per unit (price for \texttt{r3.2xlarge} per instance hour divided by 6), the Spot Fleet would launch four of these instances (24 divided by 6).

If the second launch specification provides the lowest price per unit (price for \texttt{c3.xlarge} per instance hour divided by 5), the Spot Fleet would launch five of these instances (24 divided by 5, result rounded up).

**Instance Weighting and Allocation Strategy**

Consider a Spot Fleet request with the following configuration:

- A target capacity of 30
- A launch specification with an instance type \texttt{c3.2xlarge} and a weight of 8
- A launch specification with an instance type \texttt{m3.xlarge} and a weight of 8
- A launch specification with an instance type \texttt{r3.xlarge} and a weight of 8

The Spot Fleet would launch four instances (30 divided by 8, result rounded up). With the \texttt{lowestPrice} strategy, all four instances come from the pool that provides the lowest price per unit. With the \texttt{diversified} strategy, the Spot Fleet launches one instance in each of the three pools, and the fourth instance in whichever pool provides the lowest price per unit.

**Walkthrough: Using Spot Fleet with Instance Weighting**

This walkthrough uses a fictitious company called Example Corp to illustrate the process of requesting a Spot Fleet using instance weighting.

**Objective**

Example Corp, a pharmaceutical company, wants to leverage the computational power of Amazon EC2 for screening chemical compounds that might be used to fight cancer.

**Planning**

Example Corp first reviews \texttt{Spot Best Practices}. Next, Example Corp determines the following requirements for their Spot Fleet.

**Instance Types**

Example Corp has a compute- and memory-intensive application that performs best with at least 60 GB of memory and eight virtual CPUs (vCPUs). They want to maximize these resources for the application at the lowest possible price. Example Corp decides that any of the following EC2 instance types would meet their needs:

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Memory (GiB)</th>
<th>vCPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.2xlarge</td>
<td>61</td>
<td>8</td>
</tr>
<tr>
<td>r3.4xlarge</td>
<td>122</td>
<td>16</td>
</tr>
</tbody>
</table>
Target Capacity in Units

With instance weighting, target capacity can equal a number of instances (the default) or a combination of factors such as cores (vCPUs), memory (GiBs), and storage (GBs). By considering the base for their application (60 GB of RAM and eight vCPUs) as 1 unit, Example Corp decides that 20 times this amount would meet their needs. So the company sets the target capacity of their Spot Fleet request to 20.

Instance Weights

After determining the target capacity, Example Corp calculates instance weights. To calculate the instance weight for each instance type, they determine the units of each instance type that are required to reach the target capacity as follows:

- r3.2xlarge (61.0 GB, 8 vCPUs) = 1 unit of 20
- r3.4xlarge (122.0 GB, 16 vCPUs) = 2 units of 20
- r3.8xlarge (244.0 GB, 32 vCPUs) = 4 units of 20

Therefore, Example Corp assigns instance weights of 1, 2, and 4 to the respective launch configurations in their Spot Fleet request.

Price Per Unit Hour

Example Corp uses the On-Demand price per instance hour as a starting point for their price. They could also use recent Spot prices, or a combination of the two. To calculate the price per unit hour, they divide their starting price per instance hour by the weight. For example:

<table>
<thead>
<tr>
<th>Instance type</th>
<th>On-Demand price</th>
<th>Instance weight</th>
<th>Price per unit hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.2xLarge</td>
<td>$0.7</td>
<td>1</td>
<td>$0.7</td>
</tr>
<tr>
<td>r3.4xLarge</td>
<td>$1.4</td>
<td>2</td>
<td>$0.7</td>
</tr>
<tr>
<td>r3.8xLarge</td>
<td>$2.8</td>
<td>4</td>
<td>$0.7</td>
</tr>
</tbody>
</table>

Example Corp could use a global price per unit hour of $0.7 and be competitive for all three instance types. They could also use a global price per unit hour of $0.7 and a specific price per unit hour of $0.9 in the r3.8xlarge launch specification.

Verifying Permissions

Before creating a Spot Fleet request, Example Corp verifies that it has an IAM role with the required permissions. For more information, see Spot Fleet Prerequisites (p. 255).

Creating the Request

Example Corp creates a file, config.json, with the following configuration for its Spot Fleet request:

```json
{
    "SpotPrice": "$0.70",
    "TargetCapacity": 20,
    "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
}```
"LaunchSpecifications": [  
  {  
    "ImageId": "ami-1a2b3c4d",  
    "InstanceType": "r3.2xlarge",  
    "SubnetId": "subnet-482e4972",  
    "WeightedCapacity": 1  
  },  
  {  
    "ImageId": "ami-1a2b3c4d",  
    "InstanceType": "r3.4xlarge",  
    "SubnetId": "subnet-482e4972",  
    "WeightedCapacity": 2  
  },  
  {  
    "ImageId": "ami-1a2b3c4d",  
    "InstanceType": "r3.8xlarge",  
    "SubnetId": "subnet-482e4972",  
    "SpotPrice": "0.90",  
    "WeightedCapacity": 4  
  }  
]  

Example Corp creates the Spot Fleet request using the following request-spot-fleet command:

```
aws ec2 request-spot-fleet --spot-fleet-request-config file://config.json
```

For more information, see Spot Fleet Requests (p. 254).

Fulfillment

The allocation strategy determines which Spot Instance pools your Spot Instances come from.

With the lowestPrice strategy (which is the default strategy), the Spot Instances come from the pool with the lowest price per unit at the time of fulfillment. To provide 20 units of capacity, the Spot Fleet launches either 20 r3.2xlarge instances (20 divided by 1), 10 r3.4xlarge instances (20 divided by 2), or 5 r3.8xlarge instances (20 divided by 4).

If Example Corp used the diversified strategy, the Spot Instances would come from all three pools. The Spot Fleet would launch 6 r3.2xlarge instances (which provide 6 units), 3 r3.4xlarge instances (which provide 6 units), and 2 r3.8xlarge instances (which provide 8 units), for a total of 20 units.

Spot Instance Pricing History

When you request Spot Instances, we recommend that you use the default maximum price (the On-Demand price). If you want to specify a maximum price, we recommend that you review the Spot price history before you do so. You can view the Spot price history for the last 90 days, filtering by instance type, operating system, and Availability Zone.

To view the Spot price history (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the navigation pane, choose Spot Requests.
3. If you are new to Spot Instances, you see a welcome page. Choose Get started, scroll to the bottom of the screen, and then choose Cancel.
4. Choose Pricing History. By default, the page displays a graph of the data for Linux t1.micro instances in all Availability Zones over the past day. Move your pointer over the graph to display the prices at specific times in the table below the graph.
5. (Optional) To review the Spot price history for a specific Availability Zone, select a zone from the list. You can also select a different product, instance type, or date range.

**To view the Spot price history using the command line**

You can use one of the following commands. For more information, see Accessing Amazon EC2 (p. 3).

- `describe-spot-price-history` (AWS CLI)
- `Get-EC2SpotPriceHistory` (AWS Tools for Windows PowerShell)

**Savings From Purchasing Spot Instances**

You can view the usage and savings information for Spot Instances at the per-fleet level, or for all running Spot Instances. At the per-fleet level, the usage and savings information includes all instances launched and terminated by the fleet. You can view this information from the last hour or the last three days.

The following screenshot from the Spot Requests page shows the Spot usage and savings information for a Spot Fleet.

You can view the following usage and savings information:

- **Spot Instances** – The number of Spot Instances launched and terminated by the Spot Fleet. When viewing the savings summary, the number represents all your running Spot Instances.
• **vCPU-hours** – The number of vCPU hours used across all the Spot Instances for the selected time frame.

• **Mem(GiB)-hours** – The number of GiB hours used across all the Spot Instances for the selected time frame.

• **On-Demand total** – The total amount you would've paid for the selected time frame had you launched these instances as On-Demand Instances.

• **Spot total** – The total amount to pay for the selected time frame.

• **Savings** – The percentage that you are saving by not paying the On-Demand price.

• **Average cost per vCPU-hour** – The average hourly cost of using the vCPUs across all the Spot Instances for the selected time frame, calculated as follows: \( \text{Average cost per vCPU-hour} = \frac{\text{Spot total}}{\text{vCPU-hours}} \).

• **Average cost per mem(GiB)-hour** – The average hourly cost of using the GiBs across all the Spot Instances for the selected time frame, calculated as follows: \( \text{Average cost per mem(GiB)-hour} = \frac{\text{Spot total}}{\text{Mem(GiB)-hours}} \).

• **Details table** – The different instance types (the number of instances per instance type is in parentheses) that comprise the Spot Fleet. When viewing the savings summary, these comprise all your running Spot Instances.

Savings information can only be viewed using the Amazon EC2 console.

**To view the savings information for a Spot Fleet (console)**
1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. On the navigation pane, choose **Spot Requests**.
3. Select a Spot Fleet request and choose **Savings**.
4. By default, the page displays usage and savings information for the last three days. You can choose **last hour** or the **last three days**. For Spot Fleets that were launched less than an hour ago, the page shows the estimated savings for the hour.

**To view the savings information for all running Spot Instances (console)**
1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. On the navigation pane, choose **Spot Requests**.
3. Choose **Savings Summary**.

**Spot Instance Requests**

To use Spot Instances, you create a Spot Instance request that includes the number of instances, the instance type, the Availability Zone, and the maximum price that you are willing to pay per instance hour. If your maximum price exceeds the current Spot price, Amazon EC2 fulfills your request immediately if capacity is available. Otherwise, Amazon EC2 waits until your request can be fulfilled or until you cancel the request.

The following illustration shows how Spot requests work. Notice that the action taken for a Spot Instance interruption depends on the request type (one-time or persistent) and the interruption behavior (hibernate, stop, or terminate). If the request is a persistent request, the request is opened again after your Spot Instance is interrupted.
Spot Instance Request States

A Spot Instance request can be in one of the following states:

- **open** – The request is waiting to be fulfilled.
- **active** – The request is fulfilled and has an associated Spot Instance.
- **failed** – The request has one or more bad parameters.
- **closed** – The Spot Instance was interrupted or terminated.
- **cancelled** – You canceled the request, or the request expired.

The following illustration represents the transitions between the request states. Notice that the transitions depend on the request type (one-time or persistent).

A one-time Spot Instance request remains active until Amazon EC2 launches the Spot Instance, the request expires, or you cancel the request. If the Spot price exceeds your maximum price or capacity is not available, your Spot Instance is terminated and the Spot Instance request is closed.
A persistent Spot Instance request remains active until it expires or you cancel it, even if the request is fulfilled. If the Spot price exceeds your maximum price or capacity is not available, your Spot Instance is interrupted. After your instance is interrupted, when your maximum price exceeds the Spot price or capacity becomes available again, the Spot Instance is started if stopped or resumed if hibernated. If the Spot Instance is terminated, the Spot Instance request is opened again and Amazon EC2 launches a new Spot Instance.

You can track the status of your Spot Instance requests, as well as the status of the Spot Instances launched, through the status. For more information, see Spot Request Status (p. 279).

Specifying a Duration for Your Spot Instances

Spot Instances with a specified duration (also known as Spot blocks) are designed not to be interrupted and will run continuously for the duration you select. This makes them ideal for jobs that take a finite time to complete, such as batch processing, encoding and rendering, modeling and analysis, and continuous integration.

You can specify a duration of 1, 2, 3, 4, 5, or 6 hours. The price that you pay depends on the specified duration. To view the current prices for a 1-hour duration or a 6-hour duration, see Spot Instance Prices. You can use these prices to estimate the cost of the 2, 3, 4, and 5-hour durations. When a request with a duration is fulfilled, the price for your Spot Instance is fixed, and this price remains in effect until the instance terminates. You are billed at this price for each hour or partial hour that the instance is running. A partial instance hour is billed as a full hour.

When you specify a duration in your Spot request, the duration period for each Spot Instance starts as soon as the instance receives its instance ID. The Spot Instance runs until you terminate it or the duration period ends. At the end of the duration period, Amazon EC2 marks the Spot Instance for termination and provides a Spot Instance termination notice, which gives the instance a two-minute warning before it terminates. In rare situations, Spot blocks may be interrupted due to Amazon EC2 capacity needs. In these cases, we provide a two-minute warning before we terminate an instance, and you are not charged for the terminated instances even if you used them.

To launch Spot Instances with a specified duration (console)

Select the appropriate request type. For more information, see Creating a Spot Instance Request (p. 248).

To launch Spot Instances with a specified duration (AWS CLI)

To specify a duration for your Spot Instances, include the ––block-duration-minutes option with the request-spot-instances command. For example, the following command creates a Spot request that launches Spot Instances that run for two hours:

```bash
aws ec2 request-spot-instances --instance-count 5 --block-duration-minutes 120 --type "one-time" --launch-specification file://specification.json
```

To retrieve the cost for Spot Instances with a specified duration (AWS CLI)

Use the describe-spot-instance-requests command to retrieve the fixed cost for your Spot Instances with a specified duration. The information is in the actualBlockHourlyPrice field.

Specifying a Tenancy for Your Spot Instances

You can run a Spot Instance on single-tenant hardware. Dedicated Spot Instances are physically isolated from instances that belong to other AWS accounts. For more information, see Dedicated Instances (p. 312) and the Amazon EC2 Dedicated Instances product page.

To run a Dedicated Spot Instance, do one of the following:
• Specify a tenancy of dedicated when you create the Spot Instance request. For more information, see Creating a Spot Instance Request (p. 248).

• Request a Spot Instance in a VPC with an instance tenancy of dedicated. For more information, see Creating a VPC with an Instance Tenancy of Dedicated (p. 315). You cannot request a Spot Instance with a tenancy of default if you request it in a VPC with an instance tenancy of dedicated.

The following instance types support Dedicated Spot Instances.

**Current Generation**

- c4.8xlarge
- d2.8xlarge
- i3.16xlarge
- m4.10xlarge
- m4.16xlarge
- p2.16xlarge
- r4.16xlarge
- x1.32xlarge

**Previous Generation**

- c3.8xlarge
- cc2.8xlarge
- cr1.8xlarge
- g2.8xlarge
- i2.8xlarge
- r3.8xlarge

**Service-Linked Role for Spot Instance Requests**

Amazon EC2 uses service-linked roles for the permissions that it requires to call other AWS services on your behalf. A service-linked role is a unique type of IAM role that is linked directly to an AWS service. Service-linked roles provide a secure way to delegate permissions to AWS services because only the linked service can assume a service-linked role. For more information, see Using Service-Linked Roles in the IAM User Guide.

Amazon EC2 uses the service-linked role named AWSServiceRoleForEC2Spot to launch and manage Spot Instances on your behalf.

**Permissions Granted by AWSServiceRoleForEC2Spot**

Amazon EC2 uses AWSServiceRoleForEC2Spot to complete the following actions:

- ec2:DescribeInstances – Describe Spot Instances
- ec2:StopInstances – Stop Spot Instances
- ec2:StartInstances – Start Spot Instances

**Create the Service-Linked Role**

Under most circumstances, you don’t need to manually create a service-linked role. Amazon EC2 creates the AWSServiceRoleForEC2Spot service-linked role the first time you request a Spot Instance using the console.
If you had an active Spot Instance request before October 2017, when Amazon EC2 began supporting this service-linked role, Amazon EC2 created the **AWSServiceRoleForEC2Spot** role in your AWS account. For more information, see A New Role Appeared in My Account in the IAM User Guide.

Ensure that this role exists before you use the AWS CLI or an API to request a Spot Instance. To create the role, use the IAM console as follows.

**To manually create the AWSServiceRoleForEC2Spot service-linked role**

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose **Roles**.
3. Choose **Create role**.
4. On the **Select type of trusted entity** page, choose **EC2**, **EC2 - Spot Instances**.
5. On the next page, choose **Next: Review**.
6. On the **Review** page, choose **Create role**.

If you no longer need to use Spot Instances, we recommend that you delete the **AWSServiceRoleForEC2Spot** role. After this role is deleted from your account, Amazon EC2 will create the role again if you request Spot Instances.

**Granting Access to CMKs for Use with Encrypted AMIs and EBS Snapshots**

If you specify an encrypted AMI (p. 67) or an encrypted Amazon EBS snapshot (p. 865) for your Spot Instances and you use a customer managed customer master key (CMK) for encryption, you must grant the **AWSServiceRoleForEC2Spot** role permission to use the CMK so that Amazon EC2 can launch Spot Instances on your behalf. To do this, you must add a grant to the CMK, as shown in the following procedure.

When providing permissions, grants are an alternative to key policies. For more information, see Using Grants and Using Key Policies in AWS KMS in the AWS Key Management Service Developer Guide.

**To grant the AWSServiceRoleForEC2Spot role permissions to use the CMK**

- Use the `create-grant` command to add a grant to the CMK and to specify the principal (the **AWSServiceRoleForEC2Spot** service-linked role) that is given permission to perform the operations that the grant permits. The CMK is specified by the `key-id` parameter and the ARN of the CMK. The principal is specified by the `grantee-principal` parameter and the ARN of the **AWSServiceRoleForEC2Spot** service-linked role.

The following example is formatted for legibility.

```
aws kms create-grant
--region us-east-1
--key-id arn:aws:kms:us-east-1:444455556666:key/1234abcd-12ab-34cd-56ef-1234567890ab
--grantee-principal arn:aws:iam::111122223333:role/AWSServiceRoleForEC2Spot
--operations "Decrypt" "Encrypt" "GenerateDataKey" "GenerateDataKeyWithoutPlaintext"
"CreateGrant" "DescribeKey" "ReEncryptFrom" "ReEncryptTo"
```

**Creating a Spot Instance Request**

The process for requesting a Spot Instance is similar to the process for launching an On-Demand Instance. You can't change the parameters of your Spot Instance request, including your maximum price, after you've submitted the request.

If you request multiple Spot Instances at one time, Amazon EC2 creates separate Spot Instance requests so that you can track the status of each request separately. For more information about tracking Spot Instance requests, see Spot Request Status (p. 279).
Prerequisites

Before you begin, decide on your maximum price, how many Spot Instances you'd like, and what instance type to use. To review Spot price trends, see Spot Instance Pricing History (p. 242).

To create a Spot Instance request (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the navigation pane, choose **Spot Requests**.
3. If you are new to Spot Instances, you see a welcome page; choose **Get started**. Otherwise, choose **Request Spot Instances**.
4. For **Request type**, the default is **Request**, which specifies a one-time Spot request created using a Spot Fleet. To use Spot blocks instead, choose **Reserve for duration** and select the number of hours for the job to complete.

To use **Request and Maintain**, see Creating a Spot Fleet Request (p. 259).
5. For **Target capacity**, enter the number of units to request. You can choose instances or performance characteristics that are important to your application workload, such as vCPUs, memory, and storage.
6. For **Requirements**, do the following:
   a. [Spot Fleet] (Optional) For **Launch template**, choose a launch template. The launch template must specify an Amazon Machine Image (AMI), as you cannot override the AMI using Spot Fleet if you specify a launch template.
   b. For **AMI**, choose one of the basic AMIs provided by AWS, or choose **Use custom AMI** to specify your own AMI.
   c. For **Instance type(s)**, choose **Select**. Select the instance types that have the minimum hardware specifications that you need (vCPUs, memory, and storage).
   d. For **Network**, you can select an existing VPC or create a new one.
      [Existing VPC] Select the VPC.
      [New VPC] Choose **Create new VPC** to go to the Amazon VPC console. When you are done, return to the wizard and refresh the list.
   e. (Optional) For **Availability Zones**, the default is to let AWS choose the Availability Zones for your Spot Instances. If you prefer, you can specify specific Availability Zones.
      Select one or more Availability Zones. If you have more than one subnet in an Availability Zone, select the appropriate subnet from **Subnet**. To add subnets, select **Create new subnet** to go to the Amazon VPC console. When you are done, return to the wizard and refresh the list.
   f. (Optional) To add storage, specify additional instance store volumes or EBS volumes, depending on the instance type. You can also enable Amazon EBS optimization.
   g. (Optional) By default, basic monitoring is enabled for your instances. To enable detailed monitoring, choose **Enable CloudWatch detailed monitoring**.
   h. (Optional) To run a Dedicated Spot Instance, for **Tenancy**, choose **Dedicated - run a dedicated instance**.
   i. For **Security groups**, select one or more security groups.
   j. To connect to your instances, enable **Auto-assign IPv4 Public IP**.
   k. (Optional) To connect to your instances, specify your key pair for **Key pair name**.
   l. (Optional) To launch your Spot Instances with an IAM role, for **IAM instance profile**, specify the role.
   m. (Optional) To run a start-up script, copy it to **User data**.
   n. [Spot Fleet] To add a tag, choose **Add new tag** and type the key and value for the tag. Repeat for each tag.
7. For **Spot request fulfillment**, do the following:
   a. [Spot Fleet] For **Allocation strategy**, choose the strategy that meets your needs. For more information, see Allocation Strategy for Spot Instances (p. 236).
   b. [Spot Fleet] For **Maximum price**, you can use the default maximum price (the On-Demand price) or specify the maximum price that you are willing to pay. If your maximum price is lower than the Spot price for the instance types that you selected, your Spot Instances are not launched.
   c. (Optional) To create a request that is valid only during a specific time period, edit the values for **Request valid from** and **Request valid until**.
   d. [Spot Fleet] By default, we terminate your Spot Instances when the request expires. To keep them running after your request expires, clear **Terminate instances at expiration**.

8. (Optional) To register your Spot Instances with a load balancer, choose **Receive traffic from one or more load balancers** and select one or more Classic Load Balancers or target groups.

9. (Optional) To download a copy of the launch configuration for use with the AWS CLI, choose **JSON config**.

10. Choose **Launch**.

   [Spot Fleet] The request type is **fleet**. When the request is fulfilled, requests of type **instance** are added, where the state is **active** and the status is **fulfilled**.

   [Spot block] The request type is **block** and the initial state is **open**. When the request is fulfilled, the state is **active** and the status is **fulfilled**.

**To create a Spot Instance request (AWS CLI)**

Use the following `request-spot-instances` command to create a one-time request:

```bash
aws ec2 request-spot-instances --instance-count 5 --type "one-time" --launch-specification file://specification.json
```

Use the following `request-spot-instances` command to create a persistent request:

```bash
aws ec2 request-spot-instances --instance-count 5 --type "persistent" --launch-specification file://specification.json
```

For example launch specification files to use with these commands, see Spot Request Example Launch Specifications (p. 252). If you download a launch specification file from the console, you must use the `request-spot-fleet` command instead (the console specifies a Spot request using a Spot Fleet).

Amazon EC2 launches your Spot Instance when the maximum price exceeds the Spot price and capacity is available. The Spot Instance runs until it is interrupted or you terminate it yourself. Use the following `describe-spot-instance-requests` command to monitor your Spot Instance request:

```bash
aws ec2 describe-spot-instance-requests --spot-instance-request-ids sir-08b93456
```

**Finding Running Spot Instances**

Amazon EC2 launches a Spot Instance when the maximum price exceeds the Spot price and capacity is available. A Spot Instance runs until it is interrupted or you terminate it yourself. If your maximum price is exactly equal to the Spot price, there is a chance that your Spot Instance remains running, depending on demand.

**To find running Spot Instances (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Spot Requests.

You can see both Spot Instance requests and Spot Fleet requests. If a Spot Instance request has been fulfilled, Capacity is the ID of the Spot Instance. For a Spot Fleet, Capacity indicates how much of the requested capacity has been fulfilled. To view the IDs of the instances in a Spot Fleet, choose the expand arrow, or select the fleet and choose Instances.

Note
Spot Instance requests are not tagged instantly and for a period of time may appear separate from Spot Fleet Requests (SFR).

3. Alternatively, in the navigation pane, choose Instances. In the top right corner, choose the Show/Hide icon, and then select Lifecycle. For each instance, Lifecycle is either normal, spot, or scheduled.

To find running Spot Instances (AWS CLI)

To enumerate your Spot Instances, use the describe-spot-instance-requests command with the --query option as follows:

```
aws ec2 describe-spot-instance-requests --query SpotInstanceRequests[*].{ID:InstanceId}
```

The following is example output:

```
[
  {
    "ID": "i-1234567890abcdef0"
  },
  {
    "ID": "i-0598c7d356eba48d7"
  }
]
```

Alternatively, you can enumerate your Spot Instances using the describe-instances command with the --filters option as follows:

```
aws ec2 describe-instances --filters "Name=instance-lifecycle,Values=spot"
```

Tagging Spot Instance Requests

To help categorize and manage your Spot Instance requests, you can tag them with metadata of your choice. For more information, see Tagging Your Amazon EC2 Resources (p. 964).

You can assign a tag to a Spot Instance request after you create it. The tags that you create for your Spot Instance requests only apply to the requests. These tags are not added automatically to the Spot Instance that the Spot service launches to fulfill the request. You must add tags to a Spot Instance yourself after the Spot Instance is launched.

To add a tag to your Spot Instance request or Spot Instance using the AWS CLI

Use the following create-tags command to tag your resources:

```
aws ec2 create-tags --resources sir-08b93456 i-1234567890abcdef0 --tags
   Key=purpose,Value=test
```

Canceling a Spot Instance Request

If you no longer want your Spot request, you can cancel it. You can only cancel Spot Instance requests that are open or active. Your Spot request is open when your request has not yet been fulfilled and
no instances have been launched. Your Spot request is active when your request has been fulfilled and Spot Instances have launched as a result. If your Spot request is active and has an associated running Spot Instance, canceling the request does not terminate the instance. For more information about terminating a Spot Instance, see the next section.

To cancel a Spot Instance request (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Spot Requests and select the Spot request.
3. Choose Actions, Cancel spot request.
4. (Optional) If you are finished with the associated Spot Instances, you can terminate them. In the navigation pane, choose Instances, select the instance, and then choose Actions, Instance State, Terminate.

To cancel a Spot Instance request (AWS CLI)

- Use the following cancel-spot-instance-requests command to cancel the specified Spot request:

```
aws ec2 cancel-spot-instance-requests --spot-instance-request-ids sir-08b93456
```

Terminating a Spot Instance

If your Spot request is active and has an associated running Spot Instance, canceling the request does not terminate the instance; you must terminate the running Spot Instance manually. If you terminate a running Spot Instance that was launched by a persistent Spot request, the Spot request returns to the open state so that a new Spot Instance can be launched. To cancel a persistent Spot request and terminate its Spot Instances, you must cancel the Spot request first and then terminate the Spot Instances. Otherwise, the persistent Spot request can launch a new instance. For more information about canceling a Spot Instance request, see the previous section.

To manually terminate a Spot Instance (AWS CLI)

- Use the following terminate-instances command to manually terminate Spot Instances:

```
aws ec2 terminate-instances --instance-ids i-1234567890abcdef0 i-0598c7d356eba48d7
```

Spot Request Example Launch Specifications

The following examples show launch configurations that you can use with the request-spot-instances command to create a Spot Instance request. For more information, see Creating a Spot Instance Request (p. 248).

1. Launch Spot Instances (p. 252)
2. Launch Spot Instances in the specified Availability Zone (p. 253)
3. Launch Spot Instances in the specified subnet (p. 253)
4. Launch a Dedicated Spot Instance (p. 254)

Example 1: Launch Spot Instances

The following example does not include an Availability Zone or subnet. Amazon EC2 selects an Availability Zone for you. Amazon EC2 launches the instances in the default subnet of the selected Availability Zone.
Example 2: Launch Spot Instances in the Specified Availability Zone

The following example includes an Availability Zone. Amazon EC2 launches the instances in the default subnet of the specified Availability Zone.

```json
{
  "ImageId": "ami-1a2b3c4d",
  "KeyName": "my-key-pair",
  "SecurityGroupIds": [ "sg-1a2b3c4d" ],
  "InstanceType": "m3.medium",
  "Placement": {
    "AvailabilityZone": "us-west-2a"
  },
  "IamInstanceProfile": {
    "Arn": "arn:aws:iam::123456789012:instance-profile/my-iam-role"
  }
}
```

Example 3: Launch Spot Instances in the Specified Subnet

The following example includes a subnet. Amazon EC2 launches the instances in the specified subnet. If the VPC is a nondefault VPC, the instance does not receive a public IPv4 address by default.

```json
{
  "ImageId": "ami-1a2b3c4d",
  "SecurityGroupIds": [ "sg-1a2b3c4d" ],
  "InstanceType": "m3.medium",
  "SubnetId": "subnet-1a2b3c4d",
  "IamInstanceProfile": {
    "Arn": "arn:aws:iam::123456789012:instance-profile/my-iam-role"
  }
}
```

To assign a public IPv4 address to an instance in a nondefault VPC, specify the `AssociatePublicIpAddress` field as shown in the following example. When you specify a network interface, you must include the subnet ID and security group ID using the network interface, rather than using the `SubnetId` and `SecurityGroupIds` fields shown in example 3.

```json
{
  "ImageId": "ami-1a2b3c4d",
  "KeyName": "my-key-pair",
  "InstanceType": "m3.medium",
  "NetworkInterfaces": [
    {
      "DeviceIndex": 0,
      "SubnetId": "subnet-1a2b3c4d",
      "Groups": [ "sg-1a2b3c4d" ],
      "AssociatePublicIpAddress": true
    }
  ],
  "IamInstanceProfile": {
    "Arn": "arn:aws:iam::123456789012:instance-profile/my-iam-role"
  }
}
```
Example 4: Launch a Dedicated Spot Instance

The following example requests Spot Instance with a tenancy of dedicated. A Dedicated Spot Instance must be launched in a VPC.

```
{
  "ImageId": "ami-1a2b3c4d",
  "KeyName": "my-key-pair",
  "SecurityGroupIds": [ "sg-1a2b3c4d" ],
  "InstanceType": "c3.8xlarge",
  "SubnetId": "subnet-1a2b3c4d",
  "Placement": {
    "Tenancy": "dedicated"
  }
}
```

Spot Fleet Requests

To use a Spot Fleet, you create a Spot Fleet request that includes the target capacity, an optional On-Demand portion, one or more launch specifications for the instances, and the maximum price that you are willing to pay. Amazon EC2 attempts to maintain your Spot Fleet's target capacity as Spot prices change. For more information, see How Spot Fleet Works (p. 235).

There are two types of Spot Fleet requests: request and maintain. You can create a Spot Fleet to submit a one-time request for your desired capacity, or require it to maintain a target capacity over time. Both types of requests benefit from Spot Fleet's allocation strategy.

When you make a one-time request, Spot Fleet places the required requests but does not attempt to replenish Spot Instances if capacity is diminished. If capacity is not available, Spot Fleet does not submit requests in alternative Spot pools.

To maintain a target capacity, Spot Fleet places requests to meet the target capacity and automatically replenish any interrupted instances.

It is not possible to modify the target capacity of a one-time request after it's been submitted. To change the target capacity, cancel the request and submit a new one.

A Spot Fleet request remains active until it expires or you cancel it. When you cancel a Spot Fleet request, you may specify whether canceling your Spot Fleet request terminates the Spot Instances in your Spot Fleet.

Each launch specification includes the information that Amazon EC2 needs to launch an instance, such as an AMI, instance type, subnet or Availability Zone, and one or more security groups.

Contents

- Spot Fleet Request States (p. 255)
- Spot Fleet Prerequisites (p. 255)
- Spot Fleet and IAM Users (p. 256)
- Spot Fleet Health Checks (p. 257)
- Planning a Spot Fleet Request (p. 257)
- Service-Linked Role for Spot Fleet Requests (p. 257)
- Creating a Spot Fleet Request (p. 259)
- Monitoring Your Spot Fleet (p. 262)
Spot Fleet Request States

A Spot Fleet request can be in one of the following states:

- **submitted** – The Spot Fleet request is being evaluated and Amazon EC2 is preparing to launch the target number of Spot Instances.
- **active** – The Spot Fleet has been validated and Amazon EC2 is attempting to maintain the target number of running Spot Instances. The request remains in this state until it is modified or canceled.
- **modifying** – The Spot Fleet request is being modified. The request remains in this state until the modification is fully processed or the Spot Fleet is canceled. A one-time request cannot be modified, and this state does not apply to such Spot requests.
- **cancelled_running** – The Spot Fleet is canceled and does not launch additional Spot Instances. Its existing Spot Instances continue to run until they are interrupted or terminated. The request remains in this state until all instances are interrupted or terminated.
- **cancelled_terminating** – The Spot Fleet is canceled and its Spot Instances are terminating. The request remains in this state until all instances are terminated.
- **cancelled** – The Spot Fleet is canceled and has no running Spot Instances. The Spot Fleet request is deleted two days after its instances were terminated.

The following illustration represents the transitions between the request states. If you exceed your Spot Fleet limits, the request is canceled immediately.

Spot Fleet Prerequisites

If you use the Amazon EC2 console to create a Spot Fleet, it creates a role named `aws-ec2-spot-fleet-tagging-role` that grants the Spot Fleet permission to request, launch, terminate, and tag instances on your behalf. This role is selected when you create your Spot Fleet request. If you use the AWS CLI or an API instead, you must ensure that this role exists. You can either use the Request Spot Instances wizard (the role is created when you advance to the second page of the wizard) or use the IAM console as follows.

To create the IAM role for Spot Fleet

2. In the navigation pane, choose Roles.
3. On the Select type of trusted entity page, choose AWS service, EC2, EC2 - Spot Fleet Tagging, Next: Permissions.
4. On the Attached permissions policy page, choose Next:Review.
5. On the Review page, type a name for the role (for example, aws-ec2-spot-fleet-tagging-role) and choose Create role.

Spot Fleet and IAM Users

If your IAM users will create or manage a Spot Fleet, be sure to grant them the required permissions as follows.

To grant an IAM user permissions for Spot Fleet

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Policies, Create policy.
3. On the Create policy page, choose JSON, replace the text with the following, and choose Review policy.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": ["ec2:*"],
            "Resource": "*",
        },
        {
            "Effect": "Allow",
            "Action": ["iam:ListRoles",
                        "iam:PassRole",
                        "iam:ListInstanceProfiles"],
            "Resource": "*",
        }
    ]
}
```

The ec2:* grants an IAM user permission to call all Amazon EC2 API actions. To limit the user to specific Amazon EC2 API actions, specify those actions instead.

An IAM user must have permission to call the iam:ListRoles action to enumerate existing IAM roles, the iam:PassRole action to specify the Spot Fleet role, and the iam:ListInstanceProfiles action to enumerate existing instance profiles.

(Optional) To enable an IAM user to create roles or instance profiles using the IAM console, you must also add the following actions to the policy:

- iam:AddRoleToInstanceProfile
- iam:AttachRolePolicy
- iam:CreateInstanceProfile
- iam:CreateRole
- iam:GetRole
- iam:ListPolicies
4. On the **Review policy** page, type a policy name and description and choose **Create policy**.
5. In the navigation pane, choose **Users** and select the user.
6. Choose **Permissions, Add permissions**.
7. Choose **Attach existing policies directly**. Select the policy that you created earlier and choose **Next: Review**.
8. Choose **Add permissions**.

**Spot Fleet Health Checks**

Spot Fleet checks the health status of the Spot Instances in the fleet every two minutes. The health status of an instance is either **healthy** or **unhealthy**. Spot Fleet determines the health status of an instance using the status checks provided by Amazon EC2. If the status of either the instance status check or the system status check is **impaired** for three consecutive health checks, the health status of the instance is unhealthy. Otherwise, the health status is healthy. For more information, see **Status Checks for Your Instances** (p. 535).

You can configure your Spot Fleet to replace unhealthy instances. After enabling health check replacement, an instance is replaced after its health status is reported as unhealthy. The Spot Fleet could go below its target capacity for up to a few minutes while an unhealthy instance is being replaced.

**Requirements**

- Health check replacement is supported only with Spot Fleets that maintain a target capacity, not with one-time Spot Fleets.
- You can configure your Spot Fleet to replace unhealthy instances only when you create it.
- IAM users can use health check replacement only if they have permission to call the `ec2:DescribeInstanceStatus` action.

**Planning a Spot Fleet Request**

Before you create a Spot Fleet request, review **Spot Best Practices**. Use these best practices when you plan your Spot Fleet request so that you can provision the type of instances you want at the lowest possible price. We also recommend that you do the following:

- Determine whether you want to create a Spot Fleet that submits a one-time request for the desired target capacity, or one that maintains a target capacity over time.
- Determine the instance types that meet your application requirements.
- Determine the target capacity for your Spot Fleet request. You can set the target capacity in instances or in custom units. For more information, see **Spot Fleet Instance Weighting** (p. 239).
- Determine what portion of the Spot Fleet target capacity must be On-Demand capacity. You can specify 0 for On-Demand capacity.
- Determine your price per unit, if you are using instance weighting. To calculate the price per unit, divide the price per instance hour by the number of units (or weight) that this instance represents. If you are not using instance weighting, the default price per unit is the price per instance hour.
- Review the possible options for your Spot Fleet request. For more information, see the `request-spot-fleet` command in the **AWS CLI Command Reference**. For additional examples, see **Spot Fleet Example Configurations** (p. 264).

**Service-Linked Role for Spot Fleet Requests**

Amazon EC2 uses service-linked roles for the permissions that it requires to call other AWS services on your behalf. A service-linked role is a unique type of IAM role that is linked directly to an AWS service.
Service-linked roles provide a secure way to delegate permissions to AWS services because only the linked service can assume a service-linked role. For more information, see Using Service-Linked Roles in the IAM User Guide.

Amazon EC2 uses the service-linked role named **AWSServiceRoleForEC2SpotFleet** to launch and manage Spot Instances on your behalf.

**Permissions Granted by AWSServiceRoleForEC2SpotFleet**

Amazon EC2 uses **AWSServiceRoleForEC2SpotFleet** to complete the following actions:

- ec2:RequestSpotInstances - Request Spot Instances
- ec2:TerminateInstances - Terminate Spot Instances
- ec2:DescribeImages - Describe Amazon Machine Images (AMI) for the Spot Instances
- ec2:DescribeInstanceStatus - Describe the status of the Spot Instances
- ec2:DescribeSubnets - Describe the subnets for Spot Instances
- ec2:CreateTags - Add system tags to Spot Instances

**Create the Service-Linked Role**

Under most circumstances, you don't need to manually create a service-linked role. Amazon EC2 creates the **AWSServiceRoleForEC2SpotFleet** service-linked role the first time you create a Spot Fleet using the console.

If you had an active Spot Fleet request before October 2017, when Amazon EC2 began supporting this service-linked role, Amazon EC2 created the **AWSServiceRoleForEC2SpotFleet** role in your AWS account. For more information, see A New Role Appeared in My Account in the IAM User Guide.

Ensure that this role exists before you use the AWS CLI or an API to create a Spot Fleet. To create the role, use the IAM console as follows.

**To manually create the AWSServiceRoleForEC2SpotFleet service-linked role**

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Roles.
3. Choose Create role.
4. On the Select type of trusted entity page, choose EC2, EC2 - Spot Fleet, Next: Permissions.
5. On the next page, choose Next:Review.

If you no longer need to use Spot Fleet, we recommend that you delete the **AWSServiceRoleForEC2SpotFleet** role. After this role is deleted from your account, Amazon EC2 will create the role again if you request a Spot Fleet.

**Granting Access to CMKs for Use with Encrypted AMIs and EBS Snapshots**

If you specify an encrypted AMI (p. 67) or an encrypted Amazon EBS snapshot (p. 865) in your Spot Fleet request and you use a customer managed customer master key (CMK) for encryption, you must grant the **AWSServiceRoleForEC2SpotFleet** role permission to use the CMK so that Amazon EC2 can launch Spot Instances on your behalf. To do this, you must add a grant to the CMK, as shown in the following procedure.

When providing permissions, grants are an alternative to key policies. For more information, see Using Grants and Using Key Policies in AWS KMS in the AWS Key Management Service Developer Guide.
To grant the AWSServiceRoleForEC2SpotFleet role permissions to use the CMK

- Use the `create-grant` command to add a grant to the CMK and to specify the principal (the AWSServiceRoleForEC2SpotFleet service-linked role) that is given permission to perform the operations that the grant permits. The CMK is specified by the `key-id` parameter and the ARN of the CMK. The principal is specified by the `grantee-principal` parameter and the ARN of the AWSServiceRoleForEC2SpotFleet service-linked role.

The following example is formatted for legibility.

```bash
aws kms create-grant
  --region us-east-1
  --key-id arn:aws:kms:us-east-1:444455556666:key/1234abcd-12ab-34cd-56ef-1234567890ab
  --grantee-principal arn:aws:iam::111122223333:role/AWSServiceRoleForEC2SpotFleet
  --operations "Decrypt" "Encrypt" "GenerateDataKey" "GenerateDataKeyWithoutPlaintext" "CreateGrant" "DescribeKey" "ReEncryptFrom" "ReEncryptTo"
```

Creating a Spot Fleet Request

Using the AWS Management Console, quickly create a Spot Fleet request by choosing only your application or task need and minimum compute specs. Amazon EC2 configures a fleet that best meets your needs and follows Spot best practice. For more information, see Quickly Create a Spot Fleet Request (Console) (p. 259). Otherwise, you can modify any of the default settings. For more information, see Create a Spot Fleet Request Using Defined Parameters (Console) (p. 259).

Quickly Create a Spot Fleet Request (Console)

Follow these steps to quickly create a Spot Fleet request.

To create a Spot Fleet request using the recommended settings (console)

2. If you are new to Spot, you see a welcome page; choose Get started. Otherwise, choose Request Spot Instances.
3. For Tell us your application or task need, choose Flexible workloads, Load balancing workloads, Big data workloads, or Defined duration workloads.
4. Under Configure your instances, for Minimum compute unit, choose the minimum hardware specifications (vCPUs, memory, and storage) that you need for your application or task, either as specs or as an instance type.
   - For as specs, specify the required number of vCPUs and amount of memory.
   - For as an instance type, accept the default instance type, or choose Change instance type to choose a different instance type.
5. Under Tell us how much capacity you need, for Total target capacity, specify the number of units to request for target capacity. You can choose instances or vCPUs.
6. Review the recommended Fleet request settings based on your application or task selection, and choose Launch.

Create a Spot Fleet Request Using Defined Parameters (Console)

You can create a Spot Fleet using the parameters that you define.

To create a Spot Fleet request using defined parameters (console)

2. If you are new to Spot, you see a welcome page; choose **Get started**. Otherwise, choose **Request Spot Instances**.

3. For **Tell us your application or task need**, choose **Flexible workloads**, **Load balancing workloads**, **Big data workloads**, or **Defined duration workloads**.

4. For **Configure your instances**, do the following:
   
   a. (Optional) For **Launch template**, choose a launch template. The launch template must specify an Amazon Machine Image (AMI), as you cannot override the AMI using Spot Fleet if you specify a launch template.
      
      **Important**
      
      If you intend to specify **Optional On-Demand portion**, you must choose a launch template.

   b. For **AMI**, choose one of the basic AMIs provided by AWS, or choose **Search for AMI** to use an AMI from our user community, the AWS Marketplace, or one of your own.

   c. For **Minimum compute unit**, choose the minimum hardware specifications (vCPUs, memory, and storage) that you need for your application or task, either as **specs** or as an **instance type**.
      
      • For **as specs**, specify the required number of vCPUs and amount of memory.
      
      • For **as an instance type**, accept the default instance type, or choose **Change instance type** to choose a different instance type.

   d. (Optional) For **Network**, choose an existing VPC or create a new one.
      
      [Existing VPC] Choose the VPC.

      [New VPC] Choose **Create new VPC** to go the Amazon VPC console. When you are done, return to the wizard and refresh the list.

   e. (Optional) For **Availability Zone**, let AWS choose the Availability Zones for your Spot Instances, or specify one or more Availability Zones.
      
      If you have more than one subnet in an Availability Zone, choose the appropriate subnet from **Subnet**. To add subnets, choose **Create new subnet** to go to the Amazon VPC console. When you are done, return to the wizard and refresh the list.

   f. (Optional) For **Key pair name**, choose an existing key pair or create a new one.
      
      [Existing key pair] Choose the key pair.

      [New key pair] Choose **Create new key pair** to go the Amazon VPC console. When you are done, return to the wizard and refresh the list.

5. (Optional) For **Additional configurations**, do the following:

   a. (Optional) To add storage, specify additional instance store volumes or Amazon EBS volumes, depending on the instance type.

   b. (Optional) To enable Amazon EBS optimization, for **EBS-optimized**, choose **Launch EBS-optimized instances**.

   c. (Optional) To add temporary block-level storage for your instances, for **Instance store**, choose **Attach at launch**.

   d. (Optional) By default, basic monitoring is enabled for your instances. To enable detailed monitoring, for **Monitoring**, choose **Enable CloudWatch detailed monitoring**.

   e. (Optional) To replace unhealthy instances, for **Health check**, choose **Replace unhealthy instances**. To enable this option, you must first choose **Maintain target capacity**.

   f. (Optional) To run a Dedicated Spot Instance, for **Tenancy**, choose **Dedicated - run a dedicated instance**.

   g. (Optional) For **Security groups**, choose one or more security groups or create a new one.
[Existing security group] Choose one or more security groups.

[New security group] Choose Create new security group to go the Amazon VPC console. When you are done, return to the wizard and refresh the list.

h. (Optional) To make your instances reachable from the internet, for Auto-assign IPv4 Public IP, choose Enable.

i. (Optional) To launch your Spot Instances with an IAM role, for IAM instance profile, choose the role.

j. (Optional) To run a start-up script, copy it to User data.

k. (Optional) To add a tag, choose Add new tag and enter the key and value for the tag. Repeat for each tag.

6. For Tell us how much capacity you need, do the following:

a. For Total target capacity, specify the number of units to request for target capacity. You can choose instances or vCPUs. To specify a target capacity of 0 so that you can add capacity later, choose Maintain target capacity.

b. (Optional) For Optional On-Demand portion, specify the number of On-Demand units to request. The number must be less than the Total target capacity. Amazon EC2 calculates the difference, and allocates the difference to Spot units to request.

Important
To specify an optional On-Demand portion, you must first choose a launch template.

c. (Optional) By default, the Spot service terminates Spot Instances when they are interrupted. To maintain the target capacity, choose Maintain target capacity. You can then specify that the Spot service terminates, stops, or hibernates Spot Instances when they are interrupted. To do so, choose the corresponding option from Interruption behavior.

7. For Fleet request settings, do the following:

a. Review the fleet request and fleet allocation strategy based on your application or task selection. To change the instance types or allocation strategy, clear Apply recommendations.

b. (Optional) To remove instance types, for Fleet request, choose Remove. To add instance types, choose Select instance types.

c. (Optional) For Fleet allocation strategy, choose the strategy that meets your needs. For more information, see Allocation Strategy for Spot Instances (p. 236).

8. For Additional request details, do the following:

a. Review the additional request details. To make changes, clear Apply defaults.

b. (Optional) For IAM fleet role, you can use the default role or choose a different role. To use the default role after changing the role, choose Use default role.

c. (Optional) For Maximum price, you can use the default maximum price (the On-Demand price) or specify the maximum price you are willing to pay. If your maximum price is lower than the Spot price for the instance types that you selected, your Spot Instances are not launched.

d. (Optional) To create a request that is valid only during a specific time period, edit Request valid from and Request valid until.

e. (Optional) By default, we terminate your Spot Instances when the request expires. To keep them running after your request expires, clear Terminate the instances when the request expires.

f. (Optional) To register your Spot Instances with a load balancer, choose Receive traffic from one or more load balancers and choose one or more Classic Load Balancers or target groups.

9. (Optional) To download a copy of the launch configuration for use with the AWS CLI, choose JSON config.

10. Choose Launch.
The Spot Fleet request type is `fleet`. When the request is fulfilled, requests of type `instance` are added, where the state is `active` and the status is `fulfilled`.

**To create a Spot Fleet request using the AWS CLI**

- Use the following `request-spot-fleet` command to create a Spot Fleet request:

```
aws ec2 request-spot-fleet --spot-fleet-request-config file://config.json
```

For example configuration files, see Spot Fleet Example Configurations (p. 264).

The following is example output:

```
{
    "SpotFleetRequestId": "sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE"
}
```

**Monitoring Your Spot Fleet**

The Spot Fleet launches Spot Instances when your maximum price exceeds the Spot price and capacity is available. The Spot Instances run until they are interrupted or you terminate them.

**To monitor your Spot Fleet (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Spot Requests**.
3. Select your Spot Fleet request. To see the configuration details, choose **Description**.
4. To list the Spot Instances for the Spot Fleet, choose **Instances**.
5. To view the history for the Spot Fleet, choose **History**.

**To monitor your Spot Fleet (AWS CLI)**

Use the following `describe-spot-fleet-requests` command to describe your Spot Fleet requests:

```
aws ec2 describe-spot-fleet-requests
```

Use the following `describe-spot-fleet-instances` command to describe the Spot Instances for the specified Spot Fleet:

```
aws ec2 describe-spot-fleet-instances --spot-fleet-request-id sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE
```

Use the following `describe-spot-fleet-request-history` command to describe the history for the specified Spot Fleet request:

```
aws ec2 describe-spot-fleet-request-history --spot-fleet-request-id sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --start-time 2015-05-18T00:00:00Z
```

**Modifying a Spot Fleet Request**

You can modify an active Spot Fleet request to complete the following tasks:

- Increase the target capacity and On-Demand portion
• Decrease the target capacity and On-Demand portion

**Note**
You can't modify a one-time Spot Fleet request. You can only modify a Spot Fleet request if you selected **Maintain target capacity** when you created the Spot Fleet request.

When you increase the target capacity, the Spot Fleet launches additional Spot Instances. When you increase the On-Demand portion, the Spot Fleet launches additional On-Demand Instances.

When you increase the target capacity, the Spot Fleet launches the additional Spot Instances according to the allocation strategy for its Spot Fleet request. If the allocation strategy is **lowestPrice**, the Spot Fleet launches the instances from the lowest-priced Spot Instance pool in the Spot Fleet request. If the allocation strategy is **diversified**, the Spot Fleet distributes the instances across the pools in the Spot Fleet request.

When you decrease the target capacity, the Spot Fleet cancels any open requests that exceed the new target capacity. You can request that the Spot Fleet terminate Spot Instances until the size of the fleet reaches the new target capacity. If the allocation strategy is **lowestPrice**, the Spot Fleet terminates the instances with the highest price per unit. If the allocation strategy is **diversified**, the Spot Fleet terminates instances across the pools. Alternatively, you can request that the Spot Fleet keep the fleet at its current size, but not replace any Spot Instances that are interrupted or that you terminate manually.

When a Spot Fleet terminates an instance because the target capacity was decreased, the instance receives a Spot Instance interruption notice.

**To modify a Spot Fleet request (console)**
2. Select your Spot Fleet request.
3. Choose **Actions, Modify target capacity**.
4. In **Modify target capacity**, do the following:
   a. Enter the new target capacity and On-Demand portion.
   b. (Optional) If you are decreasing the target capacity but want to keep the fleet at its current size, clear **Terminate instances**.
   c. Choose **Submit**.

**To modify a Spot Fleet request using the AWS CLI**

Use the following **modify-spot-fleet-request** command to update the target capacity of the specified Spot Fleet request:

```
aws ec2 modify-spot-fleet-request --spot-fleet-request-id sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --target-capacity 20
```

You can modify the previous command as follows to decrease the target capacity of the specified Spot Fleet without terminating any Spot Instances as a result:

```
aws ec2 modify-spot-fleet-request --spot-fleet-request-id sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --target-capacity 10 --excess-capacity-termination-policy NoTermination
```

**Canceling a Spot Fleet Request**

When you are finished using your Spot Fleet, you can cancel the Spot Fleet request. This cancels all Spot requests associated with the Spot Fleet, so that no new Spot Instances are launched for your Spot
To cancel a Spot Fleet request (console)

2. Select your Spot Fleet request.
3. Choose Actions, Cancel spot request.
4. In Cancel spot request, verify that you want to cancel the Spot Fleet. To keep the fleet at its current size, clear Terminate instances. When you are ready, choose Confirm.

To cancel a Spot Fleet request using the AWS CLI

Use the following cancel-spot-fleet-requests command to cancel the specified Spot Fleet request and terminate the instances:

```
aws ec2 cancel-spot-fleet-requests --spot-fleet-request-ids sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --terminate-instances
```

The following is example output:

```
{
    "SuccessfulFleetRequests": [
        {
            "SpotFleetRequestId": "sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE",
            "CurrentSpotFleetRequestState": "cancelled_terminating",
            "PreviousSpotFleetRequestState": "active"
        }
    ],
    "UnsuccessfulFleetRequests": []
}
```

You can modify the previous command as follows to cancel the specified Spot Fleet request without terminating the instances:

```
aws ec2 cancel-spot-fleet-requests --spot-fleet-request-ids sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --no-terminate-instances
```

The following is example output:

```
{
    "SuccessfulFleetRequests": [
        {
            "SpotFleetRequestId": "sfr-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE",
            "CurrentSpotFleetRequestState": "cancelled_running",
            "PreviousSpotFleetRequestState": "active"
        }
    ],
    "UnsuccessfulFleetRequests": []
}
```

Spot Fleet Example Configurations

The following examples show launch configurations that you can use with the request-spot-fleet command to create a Spot Fleet request. For more information, see Creating a Spot Fleet Request (p. 259).
1. Launch Spot Instances using the lowest-priced Availability Zone or subnet in the region (p. 265)
2. Launch Spot Instances using the lowest-priced Availability Zone or subnet in a specified list (p. 265)
3. Launch Spot Instances using the lowest-priced instance type in a specified list (p. 267)
4. Override the price for the request (p. 268)
5. Launch a Spot Fleet using the diversified allocation strategy (p. 269)
6. Launch a Spot Fleet using instance weighting (p. 271)
7. Launch a Spot Fleet with On-Demand capacity (p. 272)

Example 1: Launch Spot Instances Using the Lowest-Priced Availability Zone or Subnet in the Region

The following example specifies a single launch specification without an Availability Zone or subnet. The Spot Fleet launches the instances in the lowest-priced Availability Zone that has a default subnet. The price you pay does not exceed the On-Demand price.

```json
{
    "TargetCapacity": 20,
    "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
    "LaunchSpecifications": [
        {
            "ImageId": "ami-1a2b3c4d",
            "KeyName": "my-key-pair",
            "SecurityGroups": [
                {
                    "GroupId": "sg-1a2b3c4d"
                }
            ],
            "InstanceType": "m3.medium",
            "IamInstanceProfile": {
                "Arn": "arn:aws:iam::123456789012:instance-profile/my-iam-role"
            }
        }
    ]
}
```

Example 2: Launch Spot Instances Using the Lowest-Priced Availability Zone or Subnet in a Specified List

The following examples specify two launch specifications with different Availability Zones or subnets, but the same instance type and AMI.

Availability Zones

The Spot Fleet launches the instances in the default subnet of the lowest-priced Availability Zone that you specified.

```json
{
    "TargetCapacity": 20,
    "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
    "LaunchSpecifications": [
        {
            "ImageId": "ami-1a2b3c4d",
            "KeyName": "my-key-pair",
            "SecurityGroups": [
                {
                    "GroupId": "sg-1a2b3c4d"
                }
            ],
            "InstanceType": "m3.medium",
```
"Placement": {
  "AvailabilityZone": "us-west-2a, us-west-2b"
},
"IamInstanceProfile": {
  "Arn": "arn:aws:iam::123456789012:instance-profile/my-iam-role"
}
}

Subnets

You can specify default subnets or nondefault subnets, and the nondefault subnets can be from a default VPC or a nondefault VPC. The Spot service launches the instances in whichever subnet is in the lowest-priced Availability Zone.

You can't specify different subnets from the same Availability Zone in a Spot Fleet request.

{  "TargetCapacity": 20,
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [
    {  "ImageId": "ami-1a2b3c4d",
       "KeyName": "my-key-pair",
       "SecurityGroups": [
         {  "GroupId": "sg-1a2b3c4d"
       }
     ],
     "InstanceType": "m3.medium",
     "SubnetId": "subnet-a61dafcf, subnet-65ea5f08",
     "IamInstanceProfile": {
       "Arn": "arn:aws:iam::123456789012:instance-profile/my-iam-role"
     }
  ]
}

If the instances are launched in a default VPC, they receive a public IPv4 address by default. If the instances are launched in a nondefault VPC, they do not receive a public IPv4 address by default. Use a network interface in the launch specification to assign a public IPv4 address to instances launched in a nondefault VPC. When you specify a network interface, you must include the subnet ID and security group ID using the network interface.

...  {  "ImageId": "ami-1a2b3c4d",
       "KeyName": "my-key-pair",
       "InstanceType": "m3.medium",
       "NetworkInterfaces": [
         {  "DeviceIndex": 0,
            "SubnetId": "subnet-1a2b3c4d",
            "Groups": [ "sg-1a2b3c4d" ],
            "AssociatePublicIpAddress": true
         }
     ],
     "IamInstanceProfile": {
       "Arn": "arn:aws:iam::880185128111:instance-profile/my-iam-role"
     }
  }
}
Example 3: Launch Spot Instances Using the Lowest-Priced Instance Type in a Specified List

The following examples specify two launch configurations with different instance types, but the same AMI and Availability Zone or subnet. The Spot Fleet launches the instances using the specified instance type with the lowest price.

### Availability Zone

```
{
  "TargetCapacity": 20,
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [
    {
      "ImageId": "ami-1a2b3c4d",
      "SecurityGroups": [
        {
          "GroupId": "sg-1a2b3c4d"
        }
      ],
      "InstanceType": "cc2.8xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2b"
      }
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "SecurityGroups": [
        {
          "GroupId": "sg-1a2b3c4d"
        }
      ],
      "InstanceType": "r3.8xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2b"
      }
    }
  ]
}
```

### Subnet

```
{
  "TargetCapacity": 20,
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [
    {
      "ImageId": "ami-1a2b3c4d",
      "SecurityGroups": [
        {
          "GroupId": "sg-1a2b3c4d"
        }
      ],
      "InstanceType": "cc2.8xlarge",
      "SubnetId": "subnet-1a2b3c4d"
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "SecurityGroups": [
        {
          "GroupId": "sg-1a2b3c4d"
        }
      ],
      "InstanceType": "r3.8xlarge",
      "SubnetId": "subnet-1a2b3c4d"
    }
  ]
}
```
Example 4. Override the Price for the Request

We recommended that you use the default maximum price, which is the On-Demand price. If you prefer, you can specify a maximum price for the fleet request and maximum prices for individual launch specifications.

The following examples specify a maximum price for the fleet request and maximum prices for two of the three launch specifications. The maximum price for the fleet request is used for any launch specification that does not specify a maximum price. The Spot Fleet launches the instances using the instance type with the lowest price.

Availability Zone

```
{
  "SpotPrice": "1.00",
  "TargetCapacity": 30,
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [  
    {  
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c3.2xlarge",
      "Placement": {  
        "AvailabilityZone": "us-west-2b"
      },
      "SpotPrice": "0.10"
    },
    {  
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c3.4xlarge",
      "Placement": {  
        "AvailabilityZone": "us-west-2b"
      },
      "SpotPrice": "0.20"
    },
    {  
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c3.8xlarge",
      "Placement": {  
        "AvailabilityZone": "us-west-2b"
      }
    }
  ]
}
```

Subnet

```
{
  "SpotPrice": "1.00",
  "TargetCapacity": 30,
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [  
    {  
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c3.2xlarge",
      "SubnetId": "subnet-1a2b3c4d",
      "SpotPrice": "0.10"
    }
  ]
}
```
Example 5: Launch a Spot Fleet Using the Diversified Allocation Strategy

The following example uses the diversified allocation strategy. The launch specifications have different instance types but the same AMI and Availability Zone or subnet. The Spot Fleet distributes the 30 instances across the three launch specifications, such that there are 10 instances of each type. For more information, see Allocation Strategy for Spot Instances (p. 236).

Availability Zone

```json
{
  "SpotPrice": "0.70",
  "TargetCapacity": 30,
  "AllocationStrategy": "diversified",
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c4.2xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2b"
      }
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "m3.2xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2b"
      }
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "r3.2xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2b"
      }
    }
  ]
}
```

Subnet

```json
{
  "SpotPrice": "0.70",
  "TargetCapacity": 30,
  "AllocationStrategy": "diversified",
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c3.4xlarge",
      "SubnetId": "subnet-1a2b3c4d",
      "SpotPrice": "0.20"
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c3.8xlarge",
      "SubnetId": "subnet-1a2b3c4d"
    }
  ]
}
```
A best practice to increase the chance that a spot request can be fulfilled by EC2 capacity in the event of an outage in one of the Availability Zones is to diversify across AZs. For this scenario, include each AZ available to you in the launch specification. And, instead of using the same subnet each time, use three unique subnets (each mapping to a different AZ).

**Availability Zone**

```json
{
  "SpotPrice": "0.70",
  "TargetCapacity": 30,
  "AllocationStrategy": "diversified",
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c4.2xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2a"
      }
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "m3.2xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2b"
      }
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "r3.2xlarge",
      "Placement": {
        "AvailabilityZone": "us-west-2c"
      }
    }
  ]
}
```

**Subnet**

```json
{
  "SpotPrice": "0.70",
  "TargetCapacity": 30,
  "AllocationStrategy": "diversified",
  "IamFleetRole": "arn:aws:iam::123456789012:role/aws-ec2-spot-fleet-tagging-role",
  "LaunchSpecifications": [
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "c4.2xlarge",
      "SubnetId": "subnet-1a2b3c4d"
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "m3.2xlarge",
      "SubnetId": "subnet-1a2b3c4d"
    },
    {
      "ImageId": "ami-1a2b3c4d",
      "InstanceType": "r3.2xlarge",
      "SubnetId": "subnet-1a2b3c4d"
    }
  ]
}
```
Example 6: Launch a Spot Fleet Using Instance Weighting

The following examples use instance weighting, which means that the price is per unit hour instead of per instance hour. Each launch configuration lists a different instance type and a different weight. The Spot Fleet selects the instance type with the lowest price per unit hour. The Spot Fleet calculates the number of Spot Instances to launch by dividing the target capacity by the instance weight. If the result isn't an integer, the Spot Fleet rounds it up to the next integer, so that the size of your fleet is not below its target capacity.

If the r3.2xlarge request is successful, Spot provisions 4 of these instances. Divide 20 by 6 for a total of 3.33 instances, then round up to 4 instances.

If the c3.xlarge request is successful, Spot provisions 7 of these instances. Divide 20 by 3 for a total of 6.66 instances, then round up to 7 instances.

For more information, see Spot Fleet Instance Weighting (p. 239).

Availability Zone

Availability Zone

For more information, see Spot Fleet Instance Weighting (p. 239).
Example 7: Launch a Spot Fleet with On-Demand Capacity

To ensure that you always have instance capacity, you can include a request for On-Demand capacity in your Spot Fleet request. If there is capacity, the On-Demand request is always fulfilled. The balance of the target capacity is fulfilled as Spot if there is capacity and availability.

The following example specifies the desired target capacity as 10, of which 5 must be On-Demand capacity. Spot capacity is not specified; it is implied in the balance of the target capacity minus the On-Demand capacity. Amazon EC2 launches 5 capacity units as On-Demand, and 5 capacity units (10-5=5) as Spot if there is available Amazon EC2 capacity and availability.

For more information, see On-Demand in Spot Fleet (p. 236).

CloudWatch Metrics for Spot Fleet

Amazon EC2 provides Amazon CloudWatch metrics that you can use to monitor your Spot Fleet.
### Important
To ensure accuracy, we recommend that you enable detailed monitoring when using these metrics. For more information, see [Enable or Disable Detailed Monitoring for Your Instances](p. 546).

For more information about CloudWatch metrics provided by Amazon EC2, see [Monitoring Your Instances Using CloudWatch](p. 546).

### Spot Fleet Metrics

The AWS/EC2Spot namespace includes the following metrics, plus the CloudWatch metrics for the Spot Instances in your fleet. For more information, see [Instance Metrics](p. 548).

The AWS/EC2Spot namespace includes the following metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvailableInstancePoolsCount</td>
<td>The Spot Instance pools specified in the Spot Fleet request.</td>
<td>Count</td>
</tr>
<tr>
<td>BidsSubmittedForCapacity</td>
<td>The capacity for which Amazon EC2 has submitted Spot Fleet requests.</td>
<td>Count</td>
</tr>
<tr>
<td>EligibleInstancePoolCount</td>
<td>The Spot Instance pools specified in the Spot Fleet request where Amazon EC2</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>can fulfill requests. Amazon EC2 does not fulfill requests in pools where</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the maximum price you're willing to pay for Spot Instances is less than the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spot price or the Spot price is greater than the price for On-Demand</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instances.</td>
<td></td>
</tr>
<tr>
<td>FulfilledCapacity</td>
<td>The capacity that Amazon EC2 has fulfilled.</td>
<td>Count</td>
</tr>
<tr>
<td>MaxPercentCapacityAllocation</td>
<td>The maximum value of PercentCapacityAllocation across all Spot Fleet pools</td>
<td>Percent</td>
</tr>
<tr>
<td></td>
<td>specified in the Spot Fleet request.</td>
<td></td>
</tr>
<tr>
<td>PendingCapacity</td>
<td>The difference between TargetCapacity and FulfilledCapacity.</td>
<td>Count</td>
</tr>
<tr>
<td>PercentCapacityAllocation</td>
<td>The capacity allocated for the Spot Instance pool for the specified</td>
<td>Percent</td>
</tr>
<tr>
<td></td>
<td>dimensions. To get the maximum value recorded across all Spot Instance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pools, use MaxPercentCapacityAllocation.</td>
<td></td>
</tr>
<tr>
<td>TargetCapacity</td>
<td>The target capacity of the Spot Fleet request.</td>
<td>Count</td>
</tr>
<tr>
<td>TerminatingCapacity</td>
<td>The capacity that is being terminated because the provisioned capacity is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>greater than the target capacity.</td>
<td></td>
</tr>
</tbody>
</table>
### Metric

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Units: Count</td>
<td></td>
</tr>
</tbody>
</table>

If the unit of measure for a metric is Count, the most useful statistic is **Average**.

### Spot Fleet Dimensions

To filter the data for your Spot Fleet, use the following dimensions.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AvailabilityZone</td>
<td>Filter the data by Availability Zone.</td>
</tr>
<tr>
<td>FleetRequestId</td>
<td>Filter the data by Spot Fleet request.</td>
</tr>
<tr>
<td>InstanceType</td>
<td>Filter the data by instance type.</td>
</tr>
</tbody>
</table>

### View the CloudWatch Metrics for Your Spot Fleet

You can view the CloudWatch metrics for your Spot Fleet using the Amazon CloudWatch console. These metrics are displayed as monitoring graphs. These graphs show data points if the Spot Fleet is active.

Metrics are grouped first by namespace, and then by the various combinations of dimensions within each namespace. For example, you can view all Spot Fleet metrics or Spot Fleet metrics groups by Spot Fleet request ID, instance type, or Availability Zone.

**To view Spot Fleet metrics**

2. In the navigation pane, under **Metrics**, choose the **EC2 Spot** namespace.
3. (Optional) To filter the metrics by dimension, select one of the following:
   - **Fleet Request Metrics** — Group by Spot Fleet request
   - **By Availability Zone** — Group by Spot Fleet request and Availability Zone
   - **By Instance Type** — Group by Spot Fleet request and instance type
   - **By Availability Zone/Instance Type** — Group by Spot Fleet request, Availability Zone, and instance type
4. To view the data for a metric, select the check box next to the metric.
Automatic Scaling for Spot Fleet

Automatic scaling is the ability to increase or decrease the target capacity of your Spot Fleet automatically based on demand. A Spot Fleet can either launch instances (scale out) or terminate instances (scale in), within the range that you choose, in response to one or more scaling policies.

If you are using instance weighting (p. 239), keep in mind that Spot Fleet can exceed the target capacity as needed. Fulfilled capacity can be a floating-point number but target capacity must be an integer, so Spot Fleet rounds up to the next integer. You must take these behaviors into account when you look at the outcome of a scaling policy when an alarm is triggered. For example, suppose that the target capacity is 30, the fulfilled capacity is 30.1, and the scaling policy subtracts 1. When the alarm is triggered, the automatic scaling process subtracts 1 from 30.1 to get 29.1 and then rounds it up to 30, so no scaling action is taken. As another example, suppose that you selected instance weights of 2, 4, and 8, and a target capacity of 10, but no weight 2 instances were available so Spot Fleet provisioned instances of weights 4 and 8 for a fulfilled capacity of 12. If the scaling policy decreases target capacity by 20% and an alarm is triggered, the automatic scaling process subtracts 12*0.2 from 12 to get 9.6 and then rounds it up to 10, so no scaling action is taken.

You can also configure the cooldown period for a scaling policy. This is the number of seconds after a scaling activity completes where previous trigger-related scaling activities can influence future scaling events. For scale-out policies, while the cooldown period is in effect, the capacity that has been added by the previous scale-out event that initiated the cooldown is calculated as part of the desired capacity for the next scale out. The intention is to continuously (but not excessively) scale out. For scale in policies, the cooldown period is used to block subsequent scale in requests until it has expired. The intention is to scale in conservatively to protect your application’s availability. However, if another alarm triggers a scale-out policy during the cooldown period after a scale-in, automatic scaling scales out your scalable target immediately.

Spot Fleet supports the following types of automatic scaling:

- **Target tracking scaling** (p. 275) – Increase or decrease the current capacity of the fleet based on a target value for a specific metric. This is similar to the way that your thermostat maintains the temperature of your home—you select temperature and the thermostat does the rest.
- **Step scaling** (p. 276) – Increase or decrease the current capacity of the fleet based on a set of scaling adjustments, known as step adjustments, that vary based on the size of the alarm breach.
- **Scheduled scaling** (p. 278) – Increase or decrease the current capacity of the fleet based on the date and time.

For more information about configuring scaling for Spot Fleet, see the following resources:

- application-autoscaling section of the AWS CLI Command Reference
- Application Auto Scaling API Reference
- Application Auto Scaling User Guide

Scale Spot Fleet Using a Target Tracking Policy

With target tracking scaling policies, you select a metric and set a target value. Spot Fleet creates and manages the CloudWatch alarms that trigger the scaling policy and calculates the scaling adjustment based on the metric and the target value. The scaling policy adds or removes capacity as required to keep the metric at, or close to, the specified target value. In addition to keeping the metric close to the target value, a target tracking scaling policy also adjusts to the fluctuations in the metric due to a fluctuating load pattern and minimizes rapid fluctuations in the capacity of the fleet.

You can create multiple target tracking scaling policies for a Spot Fleet, provided that each of them uses a different metric. The fleet scales based on the policy that provides the largest fleet capacity. This
enables you to cover multiple scenarios and ensure that there is always enough capacity to process your application workloads.

To ensure application availability, the fleet scales out proportionally to the metric as fast as it can, but scales in more gradually.

When a Spot Fleet terminates an instance because the target capacity was decreased, the instance receives a Spot Instance interruption notice.

Do not edit or delete the CloudWatch alarms that Spot Fleet manages for a target tracking scaling policy. Spot Fleet deletes the alarms automatically when you delete the target tracking scaling policy.

**Limits**

- The Spot Fleet request must have a request type of maintain. Automatic scaling is not supported for one-time requests or Spot blocks.

**To configure a target tracking policy (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Spot Requests**.
3. Select your Spot Fleet request and choose **Auto Scaling**.
4. If automatic scaling is not configured, choose **Configure**.
5. Use **Scale capacity between** to set the minimum and maximum capacity for your fleet. Automatic scaling does not scale your fleet below the minimum capacity or above the maximum capacity.
6. For **Policy name**, type a name for the policy.
7. Choose a **Target metric**.
8. Type a **Target value** for the metric.
9. (Optional) Set **Cooldown period** to modify the default cooldown period.
10. (Optional) Select **Disable scale-in** to omit creating a scale-in policy based on the current configuration. You can create a scale-in policy using a different configuration.
11. Choose **Save**.

**To configure a target tracking policy using the AWS CLI**

1. Register the Spot Fleet request as a scalable target using the `register-scalable-target` command.
2. Create a scaling policy using the `put-scaling-policy` command.

**Scale Spot Fleet Using Step Scaling Policies**

With step scaling policies, you specify CloudWatch alarms to trigger the scaling process. For example, if you want to scale out when CPU utilization reaches a certain level, create an alarm using the `CPUUtilization` metric provided by Amazon EC2.

When you create a step scaling policy, you must specify one of the following scaling adjustment types:

- **Add** – Increase the target capacity of the fleet by a specified number of capacity units or a specified percentage of the current capacity.
- **Remove** – Decrease the target capacity of the fleet by a specified number of capacity units or a specified percentage of the current capacity.
- **Set to** – Set the target capacity of the fleet to the specified number of capacity units.
When an alarm is triggered, the automatic scaling process calculates the new target capacity using the fulfilled capacity and the scaling policy, and then updates the target capacity accordingly. For example, suppose that the target capacity and fulfilled capacity are 10 and the scaling policy adds 1. When the alarm is triggered, the automatic scaling process adds 1 to 10 to get 11, so Spot Fleet launches 1 instance.

When a Spot Fleet terminates an instance because the target capacity was decreased, the instance receives a Spot Instance interruption notice.

**Limits**

- The Spot Fleet request must have a request type of `maintain`. Automatic scaling is not supported for one-time requests or Spot blocks.

**Prerequisites**

- Consider which CloudWatch metrics are important to your application. You can create CloudWatch alarms based on metrics provided by AWS or your own custom metrics.
- For the AWS metrics that you will use in your scaling policies, enable CloudWatch metrics collection if the service that provides the metrics does not enable it by default.
- If you use the AWS Management Console to enable automatic scaling for your Spot Fleet, it creates a role named `aws-ec2-spot-fleet-autoscale-role` that grants Amazon EC2 Auto Scaling permission to describe the alarms for your policies, monitor the current capacity of the fleet, and modify the capacity of the fleet. If you configure automatic scaling using the AWS CLI or an API, you can use this role if it exists, or manually create your own role for this purpose.

**To create a role manually**

2. In the navigation pane, choose **Roles**, and then choose **Create role**.
3. For **Select type of trusted entity**, choose **AWS service**.
4. For **Choose the service that will use this role**, choose **EC2**.
5. For **Select your use case**, choose **EC2 - Spot Fleet Auto Scaling**, and then choose **Next: Permissions**.
6. For **Attached permissions policy**, the `AmazonEC2SpotFleetAutoscaleRole` policy automatically appears. Choose **Next: Tags**, and then **Next: Review**.
7. For **Review**, type a name for the role and choose **Create role**.

**To create a CloudWatch alarm**

2. In the navigation pane, choose **Alarms**.
3. Choose **Create Alarm**.
4. For **CloudWatch Metrics by Category**, choose a category. For example, choose **EC2 Spot Metrics**, **Fleet Request Metrics**.
5. Select a metric and choose **Next**.
6. For **Alarm Threshold**, type a name and description for the alarm, and set the threshold value and number of time periods for the alarm.
7. (Optional) To receive notification of a scaling event, for **Actions**, choose **New list** and type your email address. Otherwise, you can delete the notification now and add one later as needed.
8. Choose **Create Alarm**.
To configure step scaling policies for your Spot Fleet (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Spot Requests.
3. Select your Spot Fleet request and choose Auto Scaling.
4. If automatic scaling is not configured, choose Configure.
5. Use **Scale capacity between** to set the minimum and maximum capacity for your fleet. Automatic scaling does not scale your fleet below the minimum capacity or above the maximum capacity.
6. Initially, **Scaling policies** contains policies named ScaleUp and ScaleDown. You can complete these policies, or choose **Remove policy** to delete them. You can also choose **Add policy**.
7. To define a policy, do the following:
   a. For **Policy name**, type a name for the policy.
   b. For **Policy trigger**, select an existing alarm or choose **Create new alarm** to open the Amazon CloudWatch console and create an alarm.
   c. For **Modify capacity**, select a scaling adjustment type, select a number, and select a unit.
   d. (Optional) To perform step scaling, choose **Define steps**. By default, an add policy has a lower bound of -infinity and an upper bound of the alarm threshold. By default, a remove policy has a lower bound of the alarm threshold and an upper bound of +infinity. To add another step, choose **Add step**.
   e. (Optional) To modify the default value for the cooldown period, select a number from **Cooldown period**.
8. Choose **Save**.

To configure step scaling policies for your Spot Fleet using the AWS CLI

1. Register the Spot Fleet request as a scalable target using the `register-scalable-target` command.
2. Create a scaling policy using the `put-scaling-policy` command.
3. Create an alarm that triggers the scaling policy using the `put-metric-alarm` command.

Scale Spot Fleet Using Scheduled Scaling

Scaling based on a schedule enables you to scale your application in response to predictable changes in demand. To use scheduled scaling, you create **scheduled actions**, which tell Spot Fleet to perform scaling activities at specific times. When you create a scheduled action, you specify the Spot Fleet, when the scaling activity should occur, minimum capacity, and maximum capacity. You can create scheduled actions that scale one time only or that scale on a recurring schedule.

**Limits**

- The Spot Fleet request must have a request type of **maintain**. Automatic scaling is not supported for one-time requests or Spot blocks.

To create a one-time scheduled action

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Spot Requests.
3. Select your Spot Fleet request and choose Scheduled Scaling.
4. Choose **Create Scheduled Action**.
5. For **Name**, specify a name for the scheduled action.
6. Type a value for **Minimum capacity**, **Maximum capacity**, or both.
7. For **Recurrence**, choose **Once**.
8. (Optional) Choose a date and time for **Start time, End time**, or both.
9. Choose **Submit**.

**To scale on a recurring schedule**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Spot Requests**.
3. Select your Spot Fleet request and choose **Scheduled Scaling**.
4. For **Recurrence**, choose one of the predefined schedules (for example, **Every day**), or choose **Custom** and type a cron expression. For more information about the cron expressions supported by scheduled scaling, see **Cron Expressions** in the **Amazon CloudWatch Events User Guide**.
5. (Optional) Choose a date and time for **Start time, End time**, or both.
6. Choose **Submit**.

**To edit a scheduled action**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Spot Requests**.
3. Select your Spot Fleet request and choose **Scheduled Scaling**.
4. Select the scheduled action and choose **Actions, Edit**.
5. Make the needed changes and choose **Submit**.

**To delete a scheduled action**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Spot Requests**.
3. Select your Spot Fleet request and choose **Scheduled Scaling**.
4. Select the scheduled action and choose **Actions, Delete**.
5. When prompted for confirmation, choose **Delete**.

**To manage scheduled scaling using the AWS CLI**

Use the following commands:

- `put-scheduled-action`
- `describe-scheduled-actions`
- `delete-scheduled-action`

**Spot Request Status**

To help you track your Spot Instance requests and plan your use of Spot Instances, use the request status provided by Amazon EC2. For example, the request status can provide the reason why your Spot request isn't fulfilled yet, or list the constraints that are preventing the fulfillment of your Spot request.

At each step of the process—also called the Spot request lifecycle, specific events determine successive request states.

**Contents**

- Life Cycle of a Spot Request (p. 280)
Getting Request Status Information (p. 283)  
Spot Request Status Codes (p. 283)

Life Cycle of a Spot Request

The following diagram shows you the paths that your Spot request can follow throughout its lifecycle, from submission to termination. Each step is depicted as a node, and the status code for each node describes the status of the Spot request and Spot Instance.

Pending evaluation

As soon as you make a Spot Instance request, it goes into the pending-evaluation state unless one or more request parameters are not valid (bad-parameters).

<table>
<thead>
<tr>
<th>Status Code</th>
<th>Request State</th>
<th>Instance State</th>
</tr>
</thead>
<tbody>
<tr>
<td>pending-evaluation</td>
<td>open</td>
<td>n/a</td>
</tr>
<tr>
<td>bad-parameters</td>
<td>closed</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Holding

If one or more request constraints are valid but can't be met yet, or if there is not enough capacity, the request goes into a holding state waiting for the constraints to be met. The request options affect the likelihood of the request being fulfilled. For example, if you specify a maximum price below the current Spot price, your request stays in a holding state until the Spot price goes below your maximum price. If you specify an Availability Zone group, the request stays in a holding state until the Availability Zone constraint is met.

In the event of an outage of one of the Availability Zones, there is a chance that the spare EC2 capacity available for Spot Instance requests in other Availability Zones can be affected.
### Pending evaluation/fulfillment-terminal

Your Spot Instance request can go to a terminal state if you create a request that is valid only during a specific time period and this time period expires before your request reaches the pending fulfillment phase. It might also happen if you cancel the request, or if a system error occurs.

<table>
<thead>
<tr>
<th>Status Code</th>
<th>Request State</th>
<th>Instance State</th>
</tr>
</thead>
<tbody>
<tr>
<td>schedule-expired</td>
<td>cancelled</td>
<td>n/a</td>
</tr>
<tr>
<td>canceled-before-fulfillment*</td>
<td>cancelled</td>
<td>n/a</td>
</tr>
<tr>
<td>bad-parameters</td>
<td>failed</td>
<td>n/a</td>
</tr>
<tr>
<td>system-error</td>
<td>closed</td>
<td>n/a</td>
</tr>
</tbody>
</table>

* If you cancel the request.

### Pending fulfillment

When the constraints you specified (if any) are met and your maximum price is equal to or higher than the current Spot price, your Spot request goes into the pending-fulfillment state.

At this point, Amazon EC2 is getting ready to provision the instances that you requested. If the process stops at this point, it is likely to be because it was canceled by the user before a Spot Instance was launched. It may also be because an unexpected system error occurred.

<table>
<thead>
<tr>
<th>Status Code</th>
<th>Request State</th>
<th>Instance State</th>
</tr>
</thead>
<tbody>
<tr>
<td>pending-fulfillment</td>
<td>open</td>
<td>n/a</td>
</tr>
</tbody>
</table>

### Fulfilled

When all the specifications for your Spot Instances are met, your Spot request is fulfilled. Amazon EC2 launches the Spot Instances, which can take a few minutes. If a Spot Instance is hibernated or

<table>
<thead>
<tr>
<th>Status Code</th>
<th>Request State</th>
<th>Instance State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
stopped when interrupted, it remains in this state until the request can be fulfilled again or the request is
canceled.

<table>
<thead>
<tr>
<th>Status Code</th>
<th>Request State</th>
<th>Instance State</th>
</tr>
</thead>
<tbody>
<tr>
<td>fulfilled</td>
<td>active</td>
<td>pending → running</td>
</tr>
<tr>
<td>fulfilled</td>
<td>active</td>
<td>stopped → running</td>
</tr>
</tbody>
</table>

**Fulfilled-terminal**

Your Spot Instances continue to run as long as your maximum price is at or above the Spot price, there
is available capacity for your instance type, and you don’t terminate the instance. If a change in the
Spot price or available capacity requires Amazon EC2 to terminate your Spot Instances, the Spot request
goes into a terminal state. For example, if your price equals the Spot price but Spot Instances are not
available, the status code is instance-terminated-capacity-oversubscribed. A request also
goes into the terminal state if you cancel the Spot request or terminate the Spot Instances.

<table>
<thead>
<tr>
<th>Status Code</th>
<th>Request State</th>
<th>Instance State</th>
</tr>
</thead>
<tbody>
<tr>
<td>request-canceled-and-instance-running</td>
<td>cancelled</td>
<td>running</td>
</tr>
<tr>
<td>marked-for-stop</td>
<td>active</td>
<td>running</td>
</tr>
<tr>
<td>marked-for-termination</td>
<td>closed</td>
<td>running</td>
</tr>
<tr>
<td>instance-stopped-by-price</td>
<td>disabled</td>
<td>running</td>
</tr>
<tr>
<td>instance-stopped-by-user</td>
<td>disabled</td>
<td>stopped</td>
</tr>
<tr>
<td>instance-stopped-capacity-oversubscribed</td>
<td>disabled</td>
<td>stopped</td>
</tr>
<tr>
<td>instance-stopped-no-capacity</td>
<td>disabled</td>
<td>stopped</td>
</tr>
<tr>
<td>instance-terminated-by-price</td>
<td>closed (one-time), open (persistent)</td>
<td>terminated</td>
</tr>
<tr>
<td>instance-terminated-by-schedule</td>
<td>closed</td>
<td>terminated</td>
</tr>
<tr>
<td>instance-terminated-by-service</td>
<td>cancelled</td>
<td>terminated</td>
</tr>
<tr>
<td>instance-terminated-by-user †</td>
<td>closed or cancelled *</td>
<td>terminated</td>
</tr>
<tr>
<td>instance-terminated-no-capacity</td>
<td>closed (one-time), open (persistent)</td>
<td>terminated</td>
</tr>
<tr>
<td>instance-terminated-capacity-oversubscribed</td>
<td>closed (one-time), open (persistent)</td>
<td>terminated</td>
</tr>
<tr>
<td>instance-terminated-launch-group-constraint</td>
<td>closed (one-time), open (persistent)</td>
<td>terminated</td>
</tr>
</tbody>
</table>
A Spot Instance can only get to this state if a user runs the shutdown command from the instance. We do not recommend that you do this, as the Spot service might restart the instance.

* The request state is `closed` if you terminate the instance but do not cancel the request. The request state is `cancelled` if you terminate the instance and cancel the request. Even if you terminate a Spot Instance before you cancel its request, there might be a delay before Amazon EC2 detects that your Spot Instance was terminated. In this case, the request state can either be `closed` or `cancelled`.

**Persistent requests**

When your Spot Instances are terminated (either by you or Amazon EC2), if the Spot request is a persistent request, it returns to the `pending-evaluation` state and then Amazon EC2 can launch a new Spot Instance when the constraints are met.

**Getting Request Status Information**

You can get request status information using the AWS Management Console or a command line tool.

**To get request status information (console)**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Spot Requests** and select the Spot request.
3. To check the status, choose **Description, Status**.

**To get request status information using the command line**

You can use one of the following commands. For more information about these command line interfaces, see [Accessing Amazon EC2](p. 3).

- `describe-spot-instance-requests` (AWS CLI)
- `Get-EC2SpotInstanceRequest` (AWS Tools for Windows PowerShell)

**Spot Request Status Codes**

Spot request status information is composed of a status code, the update time, and a status message. Together, these help you determine the disposition of your Spot request.

The following are the Spot request status codes:

- `az-group-constraint`
  
  Amazon EC2 cannot launch all the instances you requested in the same Availability Zone.

- `bad-parameters`
  
  One or more parameters for your Spot request are not valid (for example, the AMI you specified does not exist). The status message indicates which parameter is not valid.

- `canceled-before-fulfillment`
  
  The user canceled the Spot request before it was fulfilled.

- `capacity-not-available`
  
  There is not enough capacity available for the instances that you requested.

- `capacity-oversubscribed`
  
  There is not enough capacity available for the instances that you requested.
constraint-not-fulfillable

The Spot request can't be fulfilled because one or more constraints are not valid (for example, the Availability Zone does not exist). The status message indicates which constraint is not valid.

fulfilled

The Spot request is active, and Amazon EC2 is launching your Spot Instances.

instance-stopped-by-price

Your instance was stopped because the Spot price exceeded your maximum price.

instance-stopped-by-user

Your instance was stopped because a user ran `shutdown -h` from the instance.

instance-stopped-capacity-oversubscribed

Your instance was stopped because the number of Spot requests with maximum prices equal to or higher than the Spot price exceeded the available capacity in this Spot Instance pool. The Spot price might not have changed.

instance-stopped-no-capacity

Your instance was stopped because there was no longer enough Spot capacity available for the instance.

instance-terminated-by-price

Your instance was terminated because the Spot price exceeded your maximum price. If your request is persistent, the process restarts, so your request is pending evaluation.

instance-terminated-by-schedule

Your Spot Instance was terminated at the end of its scheduled duration.

instance-terminated-by-service

Your instance was terminated from a stopped state.

instance-terminated-by-user or spot-instance-terminated-by-user

You terminated a Spot Instance that had been fulfilled, so the request state is closed (unless it's a persistent request) and the instance state is terminated.

instance-terminated-capacity-oversubscribed

Your instance was terminated because the number of Spot requests with maximum prices equal to or higher than the Spot price exceeded the available capacity in this Spot Instance pool. The Spot price might not have changed.

instance-terminated-launch-group-constraint

One or more of the instances in your launch group was terminated, so the launch group constraint is no longer fulfilled.

instance-terminated-no-capacity

Your instance was terminated because there is no longer enough Spot capacity available for the instance.

launch-group-constraint

Amazon EC2 cannot launch all the instances that you requested at the same time. All instances in a launch group are started and terminated together.
limit-exceeded

The limit on the number of EBS volumes or total volume storage was exceeded. For more information about these limits and how to request an increase, see Amazon EBS Limits in the Amazon Web Services General Reference.

marked-for-stop

The Spot Instance is marked for stopping.

marked-for-termination

The Spot Instance is marked for termination.

not-scheduled-yet

The Spot request is not evaluated until the scheduled date.

pending-evaluation

After you make a Spot Instance request, it goes into the pending-evaluation state while the system evaluates the parameters of your request.

pending-fulfillment

Amazon EC2 is trying to provision your Spot Instances.

placement-group-constraint

The Spot request can’t be fulfilled yet because a Spot Instance can’t be added to the placement group at this time.

price-too-low

The request can’t be fulfilled yet because your maximum price is below the Spot price. In this case, no instance is launched and your request remains open.

request-canceled-and-instance-running

You canceled the Spot request while the Spot Instances are still running. The request is cancelled, but the instances remain running.

schedule-expired

The Spot request expired because it was not fulfilled before the specified date.

system-error

There was an unexpected system error. If this is a recurring issue, please contact AWS Support for assistance.

Spot Instance Interruptions

Demand for Spot Instances can vary significantly from moment to moment, and the availability of Spot Instances can also vary significantly depending on how many unused EC2 instances are available. It is always possible that your Spot Instance might be interrupted. Therefore, you must ensure that your application is prepared for a Spot Instance interruption.

An On-Demand Instance specified in an EC2 Fleet or Spot Fleet cannot be interrupted.

Contents

- Reasons for Interruption (p. 286)
- Interruption Behavior (p. 286)
- Preparing for Interruptions (p. 288)
Reasons for Interruption

The following are the possible reasons that Amazon EC2 might interrupt your Spot Instances:

- **Price** – The Spot price is greater than your maximum price.
- **Capacity** – If there are not enough unused EC2 instances to meet the demand for Spot Instances, Amazon EC2 interrupts Spot Instances. The order in which the instances are interrupted is determined by Amazon EC2.
- **Constraints** – If your request includes a constraint such as a launch group or an Availability Zone group, these Spot Instances are terminated as a group when the constraint can no longer be met.

Interruption Behavior

You can specify whether Amazon EC2 should hibernate, stop, or terminate Spot Instances when they are interrupted. You can choose the interruption behavior that meets your needs. The default is to terminate Spot Instances when they are interrupted. To change the interruption behavior, choose an option from **Interruption behavior** in the console when you are creating a Spot request, or specify **InstanceInterruptionBehavior** in the launch configuration or the launch template. To change interruption behavior in the console when you are creating a Spot request, choose **Maintain target capacity**. When you select this option, **Interruption behavior** will appear and you can then specify that the Spot service terminates, stops, or hibernates Spot Instances when they are interrupted.

Stopping Interrupted Spot Instances

You can change the behavior so that Amazon EC2 stops Spot Instances when they are interrupted if the following requirements are met.

**Requirements**

- For a Spot Instance request, the type must be **persistent**. You cannot specify a launch group in the Spot Instance request.
- For an EC2 Fleet or Spot Fleet request, the type must be **maintain**.
- The root volume must be an EBS volume, not an instance store volume.

After a Spot Instance is stopped by the Spot service, only the Spot service can restart the Spot Instance, and the same launch specification must be used.

For a Spot Instance launched by a **persistent** Spot Instance request, the Spot service restarts the stopped instance when capacity is available in the same Availability Zone and for the same instance type as the stopped instance.

If instances in an EC2 Fleet or Spot Fleet are stopped and the fleet is of type **maintain**, the Spot service launches replacement instances to maintain the target capacity. The Spot service finds the best pool(s) based on the specified allocation strategy (**lowestPrice**, **diversified**, or **InstancePoolsToUseCount**); it does not prioritize the pool with the earlier stopped instances. Later, if the allocation strategy leads to a pool containing the earlier stopped instances, the Spot service restarts the stopped instances to meet the target capacity.

For example, consider a Spot Fleet with the **lowestPrice** allocation strategy. At initial launch, a c3.large pool meets the **lowestPrice** criteria for the launch specification. Later, when the c3.large
instances are interrupted, the Spot service stops the instances and replenishes capacity from another pool that fits the `lowestPrice` strategy. This time, the pool happens to be a `c4.large` pool and the Spot service launches `c4.large` instances to meet the target capacity. Similarly, Spot Fleet could move to a `c5.large` pool the next time. In each of these transitions, the Spot service does not prioritize pools with earlier stopped instances, but rather prioritizes purely on the specified allocation strategy. The `lowestPrice` strategy can lead back to pools with earlier stopped instances. For example, if instances are interrupted in the `c5.large` pool and the `lowestPrice` strategy leads it back to the `c3.large` or `c4.large` pools, the earlier stopped instances are restarted to fulfill target capacity.

While a Spot Instance is stopped, you can modify some of its instance attributes, but not the instance type. If you detach or delete an EBS volume, it is not attached when the Spot Instance is started. If you detach the root volume and the Spot service attempts to start the Spot Instance, instance start fails and the Spot service terminates the stopped instance.

You can terminate a Spot Instance while it is stopped. If you cancel a Spot request, an EC2 Fleet, or a Spot Fleet, the Spot service terminates any associated Spot Instances that are stopped.

While a Spot Instance is stopped, you are charged only for the EBS volumes, which are preserved. With EC2 Fleet and Spot Fleet, if you have many stopped instances, you can exceed the limit on the number of EBS volumes for your account.

### Hibernating Interrupted Spot Instances

You can change the behavior so that Amazon EC2 hibernates Spot Instances when they are interrupted if the following requirements are met.

#### Requirements

- For a Spot Instance request, the type must be `persistent`. You cannot specify a launch group in the Spot Instance request.
- For an EC2 Fleet or Spot Fleet request, the type must be `maintain`.
- The root volume must be an EBS volume, not an instance store volume, and it must be large enough to store the instance memory (RAM) during hibernation.
- The following instances are supported: C3, C4, C5, M4, M5, R3, and R4, with less than 100 GB of memory.
- The following operating systems are supported: Amazon Linux 2, Amazon Linux AMI, Ubuntu with an AWS-tuned Ubuntu kernel (`linux-aws`) greater than 4.4.0-1041, and Windows Server 2008 R2 and later.
- Install the hibernation agent on a supported operating system, or use one of the following AMIs, which already include the agent:
  - Amazon Linux 2
  - Amazon Linux AMI 2017.09.1 or later
  - Ubuntu Xenial 16.04 20171121 or later
  - Windows Server 2008 R2 AMI 2017.11.19 or later
  - Windows Server 2012 or Windows Server 2012 R2 AMI 2017.11.19 or later
  - Windows Server 2016 AMI 2017.11.19 or later
  - Windows Server 2019
- Start the agent. We recommend that you use user data to start the agent on instance startup. Alternatively, you could start the agent manually.

#### Recommendation

- We strongly recommend that you use an encrypted Amazon EBS volume as the root volume, because instance memory is stored on the root volume during hibernation. This ensures that the contents
of memory (RAM) are encrypted when the data is at rest on the volume and when data is moving between the instance and volume. Use one of the following three options to ensure that the root volume is an encrypted Amazon EBS volume:

- EBS “single-step” encryption: In a single run-instances API call, you can launch encrypted EBS-backed EC2 instances from an unencrypted AMI. For more information, see Using Encryption with EBS-Backed AMIs (p. 67).
- EBS encryption by default: You can enable EBS encryption by default to ensure all new EBS volumes created in your AWS account are encrypted. For more information, see Encryption by Default (p. 868).
- Encrypted AMI: You can enable EBS encryption by using an encrypted AMI to launch your instance. If your AMI does not have an encrypted root snapshot, you can copy it to a new AMI and request encryption. For more information, see Encrypt an Unencrypted Image during Copy (p. 71) and Copying an AMI (p. 75).

When a Spot Instance is hibernated by the Spot service, the EBS volumes are preserved and instance memory (RAM) is preserved on the root volume. The private IP addresses of the instance are also preserved. Instance storage volumes and public IP addresses, other than Elastic IP addresses, are not preserved. While the instance is hibernating, you are charged only for the EBS volumes. With EC2 Fleet and Spot Fleet, if you have many hibernated instances, you can exceed the limit on the number of EBS volumes for your account.

The agent prompts the operating system to hibernate when the instance receives a signal from the Spot service. If the agent is not installed, the underlying operating system doesn't support hibernation, or there isn’t enough volume space to save the instance memory, hibernation fails and the Spot service stops the instance instead.

When the Spot service hibernates a Spot Instance, you receive an interruption notice, but you do not have two minutes before the Spot Instance is interrupted. Hibernation begins immediately. While the instance is in the process of hibernating, instance health checks might fail. When the hibernation process completes, the state of the instance is stopped.

After a Spot Instance is hibernated by the Spot service, it can only be resumed by the Spot service. The Spot service resumes the instance when capacity becomes available with a Spot price that is less than your specified maximum price.

For more information, see Preparing for Instance Hibernation (p. 289).

For information about hibernating On-Demand Instances, see Hibernate Your Instance (p. 385).

Preparing for Interruptions

Here are some best practices to follow when you use Spot Instances:

- Use the default maximum price, which is the On-Demand price.
- Ensure that your instance is ready to go as soon as the request is fulfilled by using an Amazon Machine Image (AMI) that contains the required software configuration. You can also use user data to run commands at start-up.
- Store important data regularly in a place that isn't affected when the Spot Instance terminates. For example, you can use Amazon S3, Amazon EBS, or DynamoDB.
- Divide the work into small tasks (using a Grid, Hadoop, or queue-based architecture) or use checkpoints so that you can save your work frequently.
- Use Spot Instance interruption notices to monitor the status of your Spot Instances.
- While we make every effort to provide this warning as soon as possible, it is possible that your Spot Instance is terminated before the warning can be made available. Test your application to ensure that it handles an unexpected instance termination gracefully, even if you are testing for interruption
notices. You can do so by running the application using an On-Demand Instance and then terminating the On-Demand Instance yourself.

Preparing for Instance Hibernation

You must install a hibernation agent on your instance, unless you used an AMI that already includes the agent. You must run the agent on instance startup, whether the agent was included in your AMI or you installed it yourself.

The following procedure helps you prepare a Windows instance. For directions to prepare a Linux instance, see Preparing for Instance Hibernation in the Amazon EC2 User Guide for Linux Instances.

To prepare a Windows instance

1. If your AMI doesn't include the agent, download the following files to the C:\Program Files\Amazon\Hibernate folder on your Windows instance:
   - EC2HibernateAgent.exe
   - EC2HibernateAgent.ps1
   - LICENSE.txt
2. Add the following command to the user data.

   ```powershell
   <powershell>."C:\Program Files\Amazon\Hibernate\EC2HibernateAgent.exe"</powershell>
   ```

Spot Instance Interruption Notices

The best way to protect against Spot Instance interruption is to architect your application to be fault-tolerant. In addition, you can take advantage of Spot Instance interruption notices, which provide a two-minute warning before Amazon EC2 must stop or terminate your Spot Instance. We recommend that you check for these warnings every 5 seconds.

This warning is made available as a CloudWatch event and as an item in the instance metadata (p. 478) on the Spot Instance.

If you specify hibernation as the interruption behavior, you receive an interruption notice, but you do not receive a two-minute warning because the hibernation process begins immediately.

EC2 Spot Instance Interruption Notice

When Amazon EC2 is going to interrupt your Spot Instance, it emits an event two minutes prior to the actual interruption. This event can be detected by Amazon CloudWatch Events. For more information, see the Amazon CloudWatch Events User Guide.

The following is an example of the event for Spot Instance interruption. The possible values for instance-action are hibernate, stop, and terminate.

```json
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Spot Instance Interruption Warning",
  "source": "aws.ec2",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-east-2",
  "resources": ["arn:aws:ec2:us-east-2:123456789012:instance/i-1234567890abcdef0"],
  "detail": {  
```
instance-action

If your Spot Instance is marked to be stopped or terminated by the Spot service, the instance-action item is present in your instance metadata. Otherwise, it is not present. You can retrieve instance-action as follows.

```
```

The instance-action item specifies the action and the approximate time, in UTC, when the action will occur.

The following example indicates the time at which this instance will be stopped:

```
{"action": "stop", "time": "2017-09-18T08:22:00Z"}
```

The following example indicates the time at which this instance will be terminated:

```
{"action": "terminate", "time": "2017-09-18T08:22:00Z"}
```

If Amazon EC2 is not preparing to stop or terminate the instance, or if you terminated the instance yourself, instance-action is not present and you receive an HTTP 404 error.

termination-time

This item is maintained for backward compatibility; you should use instance-action instead.

If your Spot Instance is marked for termination by the Spot service, the termination-time item is present in your instance metadata. Otherwise, it is not present. You can retrieve termination-time as follows.

```
```

The termination-time item specifies the approximate time in UTC when the instance receives the shutdown signal. For example:

```
2015-01-05T18:02:00Z
```

If Amazon EC2 is not preparing to terminate the instance, or if you terminated the Spot Instance yourself, the termination-time item is either not present (so you receive an HTTP 404 error) or contains a value that is not a time value.

If Amazon EC2 fails to terminate the instance, the request status is set to fulfilled. The termination-time value remains in the instance metadata with the original approximate time, which is now in the past.

Billing for Interrupted Spot Instances

When a Spot Instance (not in a Spot block) is interrupted, you’re charged as follows.
### Who interrupts the Spot Instance

<table>
<thead>
<tr>
<th>Who interrupts the Spot Instance</th>
<th>Operating system</th>
<th>Interrupted in the first hour</th>
<th>Interrupted in any hour after the first hour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>If you</strong> interrupt the Spot Instance</td>
<td>Linux (excluding RHEL and SUSE)</td>
<td>Charged for the seconds used</td>
<td>Charged for the seconds used</td>
</tr>
<tr>
<td></td>
<td>Windows, RHEL, SUSE</td>
<td>Charged for the full hour even if you used a partial hour</td>
<td>Charged for the full hours used, and charged a full hour for the interrupted partial hour</td>
</tr>
<tr>
<td><strong>If Amazon EC2</strong> interrupts the Spot Instance</td>
<td>Linux (excluding RHEL and SUSE)</td>
<td>No charge</td>
<td>Charged for the seconds used</td>
</tr>
<tr>
<td></td>
<td>Windows, RHEL, SUSE</td>
<td>No charge</td>
<td>Charged for the full hours used, but no charge for the interrupted partial hour</td>
</tr>
</tbody>
</table>

When a Spot Instance in a Spot block is interrupted, you’re charged as follows.

<table>
<thead>
<tr>
<th>Who interrupts the Spot Instance</th>
<th>Operating system</th>
<th>Interrupted in the first hour</th>
<th>Interrupted in any hour after the first hour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>If you</strong> interrupt the Spot Instance</td>
<td>Linux (excluding RHEL and SUSE)</td>
<td>Charged for the seconds used</td>
<td>Charged for the seconds used</td>
</tr>
<tr>
<td></td>
<td>Windows, RHEL, SUSE</td>
<td>Charged for the full hour even if you used a partial hour</td>
<td>Charged for the full hours used, and charged a full hour for the interrupted partial hour</td>
</tr>
<tr>
<td><strong>If Amazon EC2</strong> interrupts the Spot Instance</td>
<td>Linux (excluding RHEL and SUSE)</td>
<td>No charge</td>
<td>No charge</td>
</tr>
<tr>
<td></td>
<td>Windows, RHEL, SUSE</td>
<td>No charge</td>
<td>No charge</td>
</tr>
</tbody>
</table>

### Spot Instance Data Feed

To help you understand the charges for your Spot Instances, Amazon EC2 provides a data feed that describes your Spot Instance usage and pricing. This data feed is sent to an Amazon S3 bucket that you specify when you subscribe to the data feed.

Data feed files arrive in your bucket typically once an hour, and each hour of usage is typically covered in a single data file. These files are compressed (gzip) before they are delivered to your bucket. Amazon EC2 can write multiple files for a given hour of usage where files are large (for example, when file contents for the hour exceed 50 MB before compression).

**Note**

If you don’t have a Spot Instance running during a certain hour, you don’t receive a data feed file for that hour.

### Contents
- Data Feed File Name and Format (p. 292)
Data Feed File Name and Format

The Spot Instance data feed file name uses the following format (with the date and hour in UTC):

```
bucket-name.s3.amazonaws.com/{optional prefix}/aws-account-id.YYYY-MM-DD-HH.n.unique-id.gz
```

For example, if your bucket name is `myawsbucket` and your prefix is `myprefix`, your file names are similar to the following:

```
myawsbucket.s3.amazonaws.com/myprefix/111122233333.2014-03-17-20.001.pwBdGTJG.gz
```

The Spot Instance data feed files are tab-delimited. Each line in the data file corresponds to one instance hour and contains the fields listed in the following table.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timestamp</td>
<td>The timestamp used to determine the price charged for this instance usage.</td>
</tr>
<tr>
<td>UsageType</td>
<td>The type of usage and instance type being charged for. For <code>m1.small</code> Spot Instances, this field is set to <code>SpotUsage</code>. For all other instance types, this field is set to <code>SpotUsage:{instance-type}</code>. For example, <code>SpotUsage:c1.medium</code>.</td>
</tr>
<tr>
<td>Operation</td>
<td>The product being charged for. For Linux Spot Instances, this field is set to <code>RunInstances</code>. For Windows Spot Instances, this field is set to <code>RunInstances:0002</code>. Spot usage is grouped according to Availability Zone.</td>
</tr>
<tr>
<td>InstanceID</td>
<td>The ID of the Spot Instance that generated this instance usage.</td>
</tr>
<tr>
<td>MyBidID</td>
<td>The ID for the Spot Instance request that generated this instance usage.</td>
</tr>
<tr>
<td>MyMaxPrice</td>
<td>The maximum price specified for this Spot Instance request.</td>
</tr>
<tr>
<td>MarketPrice</td>
<td>The Spot price at the time specified in the <code>Timestamp</code> field.</td>
</tr>
<tr>
<td>Charge</td>
<td>The price charged for this instance usage.</td>
</tr>
<tr>
<td>Version</td>
<td>The version included in the data feed file name for this record.</td>
</tr>
</tbody>
</table>

Amazon S3 Bucket Requirements

When you subscribe to the data feed, you must specify an Amazon S3 bucket to store the data feed files. Before you choose an Amazon S3 bucket for the data feed, consider the following:

- You must have **FULL_CONTROL** permission to the bucket, which includes permission for the `s3:GetBucketAcl` and `s3:PutBucketAcl` actions.

  If you’re the bucket owner, you have this permission by default. Otherwise, the bucket owner must grant your AWS account this permission.

- When you subscribe to a data feed, these permissions are used to update the bucket ACL to give the AWS data feed account **FULL_CONTROL** permission. The AWS data feed account writes data feed files
to the bucket. If your account doesn't have the required permissions, the data feed files cannot be
written to the bucket.

Note
If you update the ACL and remove the permissions for the AWS data feed account, the data
feed files cannot be written to the bucket. You must resubscribe to the data feed to receive
the data feed files.

• Each data feed file has its own ACL (separate from the ACL for the bucket). The bucket owner
  has FULL_CONTROL permission to the data files. The AWS data feed account has read and write
  permissions.
• If you delete your data feed subscription, Amazon EC2 doesn't remove the read and write permissions
  for the AWS data feed account on either the bucket or the data files. You must remove these
  permissions yourself.

Subscribing to Your Spot Instance Data Feed

To subscribe to your data feed, use the following create-spot-datafeed-subscription command:

```
aws ec2 create-spot-datafeed-subscription --bucket myawsbucket [--prefix myprefix]
```

The following is example output:

```
{
    "SpotDatafeedSubscription": {
        "OwnerId": "111122223333",
        "Prefix": "myprefix",
        "Bucket": "myawsbucket",
        "State": "Active"
    }
}
```

Deleting Your Spot Instance Data Feed

To delete your data feed, use the following delete-spot-datafeed-subscription command:

```
aws ec2 delete-spot-datafeed-subscription
```

Spot Instance Limits

Spot Instance requests are subject to the following limits:

Limits
• Spot Request Limits (p. 293)
• Spot Fleet Limits (p. 294)
• T3 Instances (p. 294)
• T2 Instances (p. 294)

Spot Request Limits

By default, there is an account limit of 20 Spot Instances per Region. If you terminate your Spot Instance
but do not cancel the request, the request counts against this limit until Amazon EC2 detects the
termination and closes the request.
Spot Instance limits are dynamic. When your account is new, your limit might be lower than 20 to start, but can increase over time. In addition, your account might have limits on specific Spot Instance types. If you submit a Spot Instance request and you receive the error Max spot instance count exceeded, you can complete the AWS Support Center Create case form to request a Spot Instance limit increase. For Limit type, choose EC2 Spot Instances. For more information, see Amazon EC2 Service Limits (p. 974).

Spot Fleet Limits

The usual Amazon EC2 limits apply to instances launched by a Spot Fleet or an EC2 Fleet, such as Spot request price limits, instance limits, and volume limits. In addition, the following limits apply:

- The number of active Spot Fleets and EC2 Fleets per Region: 1,000*
- The number of launch specifications per fleet: 50*
- The size of the user data in a launch specification: 16 KB*
- The target capacity per Spot Fleet or EC2 Fleet: 10,000
- The target capacity across all Spot Fleets and EC2 Fleets in a Region: 100,000
- A Spot Fleet request or an EC2 Fleet request can't span Regions.
- A Spot Fleet request or an EC2 Fleet request can't span different subnets from the same Availability Zone.

If you need more than the default limits for target capacity, complete the AWS Support Center Create case form to request a limit increase. For Limit type, choose EC2 Fleet, choose a Region, and then choose Target Fleet Capacity per Fleet (in units) or Target Fleet Capacity per Region (in units), or both.

* These are hard limits. You cannot request a limit increase for these limits.

T3 Instances

If you plan to use your T3 Spot Instances immediately and for a short duration, with no idle time for accruing CPU credits, we recommend that you launch your T3 Spot Instances in standard (p. 142) mode to avoid paying higher costs.

If you launch your T3 Spot Instances in unlimited (p. 135) mode and burst CPU immediately, you'll spend surplus credits for bursting. If you use the instance for a short duration, your instance doesn't have time to accrue CPU credits to pay down the surplus credits, and you are charged for the surplus credits when you terminate your instance.

Unlimited mode for T3 Spot Instances is suitable only if the instance runs for long enough to accrue CPU credits for bursting. Otherwise, paying for surplus credits makes T3 Spot Instances more expensive than M5 or C5 instances.

T2 Instances

Launch credits are meant to provide a productive initial launch experience for T2 instances by providing sufficient compute resources to configure the instance. Repeated launches of T2 instances to access new launch credits is not permitted. If you require sustained CPU, you can earn credits (by idling over some period), use T2 Unlimited (p. 135), or use an instance type with dedicated CPU (for example, c4.large).

Dedicated Hosts

An Amazon EC2 Dedicated Host is a physical server with EC2 instance capacity fully dedicated to your use. Dedicated Hosts allow you to use your existing per-socket, per-core, or per-VM software licenses, including Windows Server, Microsoft SQL Server, SUSE, Linux Enterprise Server, and so on.
Differences between Dedicated Hosts and Dedicated Instances

Dedicated Hosts and Dedicated Instances can both be used to launch Amazon EC2 instances onto physical servers that are dedicated for your use.

There are no performance, security, or physical differences between Dedicated Instances and instances on Dedicated Hosts. The following table highlights some of the key differences between Dedicated Hosts and Dedicated Instances:

<table>
<thead>
<tr>
<th></th>
<th>Dedicated Host</th>
<th>Dedicated Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Billing</td>
<td>Per-host billing</td>
<td>Per-instance billing</td>
</tr>
<tr>
<td>Visibility of sockets, cores, and host ID</td>
<td>Provides visibility of the number of sockets and physical cores</td>
<td>No visibility</td>
</tr>
<tr>
<td>Host and instance affinity</td>
<td>Allows you to consistently deploy your instances to the same physical server over time</td>
<td>Not supported</td>
</tr>
<tr>
<td>Targeted instance placement</td>
<td>Provides additional visibility and control over how instances are placed on a physical server</td>
<td>Not supported</td>
</tr>
<tr>
<td>Automatic instance recovery</td>
<td>Supported. For more information, see Host Recovery (p. 307).</td>
<td>Supported</td>
</tr>
<tr>
<td>Bring Your Own License (BYOL)</td>
<td>Supported</td>
<td>Not supported</td>
</tr>
</tbody>
</table>

Bring Your Own License

Dedicated Hosts allow you to use your existing per-socket, per-core, or per-VM software licenses. When you bring your own license, you are responsible for managing your own licenses. However, Amazon EC2 has features that help you maintain license compliance, such as instance affinity and targeted placement.

These are the general steps to follow in order to bring your own volume licensed machine image into Amazon EC2.

1. Verify that the license terms controlling the use of your machine images allow usage in a virtualized cloud environment. For more information about Microsoft Licensing, see Amazon Web Services and Microsoft Licensing.
2. After you have verified that your machine image can be used within Amazon EC2, import it using VM Import/Export. For information about how to import your machine image, see the VM Import/Export User Guide.

3. After you import your machine image, you can launch instances from it onto active Dedicated Hosts in your account.

4. When you run these instances, depending on the operating system, you may be required to activate these instances against your own KMS server (for example, Windows Server or Windows SQL Server). You cannot activate your imported Windows AMI against the Amazon Windows KMS server.

**Note**
To track how your images are used in AWS, enable host recording in AWS Config. You can use AWS Config to record configuration changes to a Dedicated Host and use the output as a data source for license reporting. For more information, see Tracking Configuration Changes (p. 311).

### Dedicated Host Instance Capacity

Dedicated Hosts are configured to support a single instance type and size capacity. The number of instances you can launch onto a Dedicated Host depends on the instance type that the Dedicated Host is configured to support. For example, if you allocated a c3.xlarge Dedicated Host, you'd have the right to launch up to eight c3.xlarge instances on the Dedicated Host. To determine the number of instance type sizes that you can run on a particular Dedicated Host, see Amazon EC2 Dedicated Hosts Pricing.

### Dedicated Hosts Limitations and Restrictions

Before you allocate Dedicated Hosts, take note of the following limitations and restrictions:

- RHEL, SUSE Linux, and Windows AMIs (whether offered by AWS or on the AWS Marketplace) cannot be used with Dedicated Hosts.
- Up to two On-Demand Dedicated Hosts per instance family, per Region can be allocated. It is possible to request a limit increase: Request to Raise Allocation Limit on Amazon EC2 Dedicated Hosts.
- The instances that run on a Dedicated Host can only be launched in a VPC.
- Host limits are independent from instance limits. Instances that you are running on Dedicated Hosts do not count towards your instance limits.
- Auto Scaling groups are not supported.
- Amazon RDS instances are not supported.
- The AWS Free Usage tier is not available for Dedicated Hosts.
- Instance placement control refers to managing instance launches onto Dedicated Hosts. Placement groups are not supported for Dedicated Hosts.

### Pricing and Billing

#### On-Demand Dedicated Hosts

On-Demand billing is automatically activated when you allocate a Dedicated Host to your account.

The On-Demand price for a Dedicated Host varies by instance family and Region. You are charged an hourly rate for the Dedicated Host, regardless of the quantity or the size of instances that you choose to launch on it. In other words, you are charged for the entire Dedicated Host, and not the individual instances that you choose to run on it. For more information about On-Demand pricing, see Amazon EC2 Dedicated Hosts On-Demand Pricing.

You can release an On-Demand Dedicated Host at any time to stop accruing charges for it. For information about releasing a Dedicated Host, see Releasing Dedicated Hosts (p. 304).
Dedicated Host Reservations

Dedicated Host Reservations provide a billing discount compared to running On-Demand Dedicated Hosts. Reservations are available in three payment options:

- **No Upfront**—No Upfront Reservations provide you with a discount on your Dedicated Host usage over a term and do not require an upfront payment. Available for a one-year term only.
- **Partial Upfront**—A portion of the reservation must be paid upfront and the remaining hours in the term are billed at a discounted rate. Available in one-year and three-year terms.
- **All Upfront**—Provides the lowest effective price. Available in one-year and three-year terms and covers the entire cost of the term upfront, with no additional future charges.

You must have active Dedicated Hosts in your account before you can purchase reservations. Each reservation covers a single, specific Dedicated Host in your account. Reservations are applied to the instance family on the host, not the instance size. If you have three Dedicated Hosts with different instances sizes (m4.xlarge, m4.medium, and m4.large) you can associate a single m4 reservation with all those Dedicated Hosts. The instance family and Region of the reservation must match that of the Dedicated Hosts you want to associate it with.

When a reservation is associated with a Dedicated Host, the Dedicated Host can't be released until the reservation's term is over.

For more information about reservation pricing, see Amazon EC2 Dedicated Hosts Pricing.

Working with Dedicated Hosts

To use a Dedicated Host, you first allocate hosts for use in your account. You then launch instances onto the hosts by specifying host tenancy for the instance. You must select a specific host for the instance to launch on to, or you can allow it to launch on to any host that has auto-placement enabled and matches its instance type. When an instance is stopped and restarted, the Host affinity setting determines whether it's restarted on the same, or a different, host.

If you no longer need an On-Demand host, you can stop the instances running on the host, direct them to launch on a different host, and then release the host.

Contents

- Understanding Auto-Placement and Affinity (p. 297)
- Allocating Dedicated Hosts (p. 298)
- Launching Instances onto Dedicated Hosts (p. 299)
- Modifying Dedicated Host Auto-Placement (p. 300)
- Modifying Instance Tenancy and Affinity (p. 301)
- Viewing Dedicated Hosts (p. 302)
- Tagging Dedicated Hosts (p. 302)
- Monitoring Dedicated Hosts (p. 303)
- Releasing Dedicated Hosts (p. 304)
- Purchasing Dedicated Host Reservations (p. 305)
- Viewing Dedicated Host Reservations (p. 306)
- Tagging Dedicated Host Reservations (p. 307)

Understanding Auto-Placement and Affinity

Placement control happens on both the instance level and host level.
Auto-Placement

Auto-placement allows you to manage whether instances that you launch are launched onto a specific host, or onto any available host that has matching configurations. Auto-placement must be configured at the host level.

When a Dedicated Host's auto-placement is disabled, it only accepts Host tenancy instance launches that specify its unique host ID. This is the default setting for new Dedicated Hosts.

When a Dedicated Host's auto-placement is enabled, it accepts any untargeted instance launches that match its instance type configuration.

When launching an instance, you need to configure its tenancy. Launching an instance onto a Dedicated Host without providing a specific HostId, enables it to launch on any Dedicated Host that has auto-placement enabled and matches its instance type.

Host Affinity

Host Affinity is configured at the instance level. It establishes a launch relationship between an instance and a Dedicated Host.

When affinity is set to Host, an instance launched onto a specific host always restarts on the same host if stopped. This applies to both targeted and untargeted launches.

When affinity is set to Off, and you stop and restart the instance, it can be restarted on any available host. However, it tries to launch back onto the last Dedicated Host on which it ran (on a best-effort basis).

Allocating Dedicated Hosts

To begin using Dedicated Hosts, they must be allocated to your account. You can allocate Dedicated Hosts to your account using the Amazon EC2 console or the command line tools.

To allocate Dedicated Hosts using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Dedicated Hosts, Allocate Dedicated Host.
3. Configure the following Dedicated Host options:
   a. **Instance type**—The type of instance that you want to launch on the Dedicated Host.
   b. **Availability Zone**—The Availability Zone in which the Dedicated Host is located.
   c. To allow the Dedicated Host to accept untargeted instance launches that match its instance type, for **Instance auto-placement**, choose Enable.
   
   For more information about auto-placement, see Understanding Auto-Placement and Affinity (p. 297).
   d. To enable host recovery for the Dedicated Host, for Host recovery choose Enable. For more information, see Host Recovery (p. 307).
   e. **Quantity**—The number of Dedicated Hosts to allocate with these options.
4. (Optional) Choose Add Tag and enter a tag key and a tag value.
5. Choose Allocate host.

To allocate Dedicated Hosts using the command line tools

Use one of the following commands. The following commands allocate a Dedicated Host that supports untargeted m4.large instance launches in the eu-west-1a Availability Zone, enable host recovery, and apply a tag with a key of purpose and a value of production.
• **allocate-hosts** (AWS CLI)

```bash
aws ec2 allocate-hosts --instance-type "m4.large" --availability-zone "eu-west-1a" --auto-placement "off" --host-recovery "on" --quantity 1 --tag-specifications 'ResourceType=dedicated-host,Tags=[{Key=purpose,Value=production}]'
```

• **New-EC2Host** (AWS Tools for Windows PowerShell)

The `TagSpecification` parameter used to tag a Dedicated Host on creation requires an object that specifies the type of resource to be tagged, the tag key, and the tag value. The following commands create the required object.

```powershell
PS C:\> $tag = @{ Key="purpose"; Value="production" }
PS C:\> $tagspec = new-object Amazon.EC2.Model.TagSpecification
PS C:\> $tagspec.ResourceType = "dedicated-host"
PS C:\> $tagspec.Tags.Add($tag)
```

The following command allocates the Dedicated Host and applies the tag specified in the `$tagspec` object.

```powershell
PS C:\> New-EC2Host -InstanceType m4.large -AvailabilityZone eu-west-1a -AutoPlacement Off -HostRecovery On -Quantity 1 -TagSpecification $tagspec
```

The Dedicated Host capacity is made available in your account immediately.

If you launch instances with host tenancy but do not have any active Dedicated Host in your account, you receive an error and the instance launch fails.

### Launching Instances onto Dedicated Hosts

After you have allocated a Dedicated Host, you can launch instances onto it. You cannot launch instances with host tenancy if you do not have active Dedicated Hosts with enough available capacity for the instance type that you are launching.

**Note**

The instances launched onto Dedicated Hosts can only be launched in a VPC. For more information, see [*Introduction to VPC*](#).

Before you launch your instances, take note of the limitations. For more information, see [*Dedicated Hosts Limitations and Restrictions*](#) (p. 296).

To launch an instance onto a specific Dedicated Host from the Dedicated Hosts page

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. Choose **Dedicated Hosts** in the navigation pane.
3. On the **Dedicated Hosts** page, select a host choose **Actions, Launch Instance(s) onto Host**.
4. Select an AMI from the list. Windows, SUSE, and RHEL AMIs provided by Amazon EC2 can't be used with Dedicated Hosts.
5. On the **Choose an Instance Type** page, keep the instance type that is selected by default, and then choose **Next: Configure Instance Details**.

   The instance type is determined by the host that you selected.

6. On the **Configure Instance Details** page, configure the instance settings to suit your needs, and then for **Affinity**, choose one of the following options:

   • **Off**—The instance launches onto the specified host, but it is not guaranteed to restart on the same Dedicated Host if stopped.
• **Host**—If stopped, the instance always restarts on this specific host.

For more information about Affinity, see *Understanding Auto-Placement and Affinity (p. 297).*

**Note**
The Tenancy and Host options are pre-configured based on the host that you selected.

7. Choose **Review and Launch**.
8. On the **Review Instance Launch** page, choose **Launch**.
9. When prompted, select an existing key pair or create a new one, and then choose **Launch Instances**.

### To launch an instance onto a Dedicated Host using the Launch Instance wizard

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**, **Launch Instance**.
3. Select an AMI from the list. Windows, SUSE, and RHEL AMIs provided by Amazon EC2 can't be used with Dedicated Hosts.
4. Select the type of instance to launch and choose **Next: Configure Instance Details**.
5. On the **Configure Instance Details** page, configure the instance settings to suit your needs, and then configure the following Dedicated Host-specific settings:
   
   • **Tenancy**—Choose **Dedicated Host - Launch this instance on a Dedicated Host**.
   
   • **Host**—Choose either **Use auto-placement** to launch the instance on any Dedicated Host that has auto-placement enabled, or select a specific Dedicated Host in the list. If Dedicated Hosts does not support the selected instance type, it is disabled in the list.
   
   • **Affinity**—Choose one of the following options:
     • **Off**—The instance launches onto the specified host, but it is not guaranteed to restart on it if stopped.
     • **Host**—If stopped, the instance always restarts on the specified host.

For more information, see *Understanding Auto-Placement and Affinity (p. 297).*

**Note**
If you are unable to see these settings, check that you have selected a VPC in the **Network** menu.

6. Choose **Review and Launch**.
7. On the **Review Instance Launch** page, choose **Launch**.
8. When prompted, select an existing key pair or create a new one, and then choose **Launch Instances**.

### To launch an instance onto a Dedicated Host using the command line tools

Use one of the following commands and specify the instance affinity, tenancy, and host in the **Placement** request parameter:

- **run-instances** (AWS CLI)
- **New-EC2Instance** (AWS Tools for Windows PowerShell)

### Modifying Dedicated Host Auto-Placement

You can modify a Dedicated Host's auto-placement settings after you have allocated it to your AWS account.
To modify a Dedicated Host's auto-placement using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Dedicated Hosts in the navigation pane.
3. On the Dedicated Hosts page, select a host and choose Actions, Modify Auto-Placement.
4. On the Modify Auto-placement window, for Allow instance auto-placement, choose Yes to enable auto-placement, or choose No to disable auto-placement. For more information, see Understanding Auto-Placement and Affinity (p. 297).
5. Choose Save.

To modify a Dedicated Host's auto-placement using the command line tools

Use one of the following commands. The following examples enable auto-placement for the specified Dedicated Host.

- modify-hosts (AWS CLI)

```bash
aws ec2 modify-hosts --auto-placement on --host-ids h-012a3456b7890cdef
```

- Edit-EC2Host (AWS Tools for Windows PowerShell)

```powershell
PS C:\> Edit-EC2Host --AutoPlacement 1 --HostId h-012a3456b7890cdef
```

Modifying Instance Tenancy and Affinity

You can change the tenancy of an instance from dedicated to host, or from host to dedicated, after you have launched it.

To modify instance tenancy and affinity using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Instances, and select the instance to modify.
3. Choose Actions, Instance State, and Stop.
4. Open the context (right-click) menu on the instance and choose Instance Settings, Modify Instance Placement.
5. On the Modify Instance Placement page, configure the following:

   - Tenancy—Choose one of the following:
     - Run a dedicated hardware instance—Launches the instance as a Dedicated Instance. For more information, see Dedicated Instances (p. 312).
     - Launch the instance on a Dedicated Host—Launches the instance onto a Dedicated Host with configurable affinity.
   - Affinity—Choose one of the following:
     - This instance can run on any one of my hosts—The instance launches onto any available Dedicated Host in your account that supports its instance type.
     - This instance can only run on the selected host—The instance is only able to run on the Dedicated Host selected for Target Host.
   - Target Host—Select the Dedicated Host that the instance must run on. If no target host is listed, you may not have available, compatible Dedicated Hosts in your account.

For more information, see Understanding Auto-Placement and Affinity (p. 297).
6. Choose **Save**.

### To modify instance tenancy and affinity using the command line tools

Use one of the following commands. The following examples change the specified instance's affinity from **default** to **host**, and specify the Dedicated Host that the instance has affinity with.

- **modify-instance-placement** (AWS CLI)
  
  ```bash
  aws ec2 modify-instance-placement --instance-id i-1234567890abcdef0 --affinity host --host-id h-012a3456b7890cdef
  ```

- **Edit-EC2InstancePlacement** (AWS Tools for Windows PowerShell)
  
  ```powershell
  PS C:\> Edit-EC2InstancePlacement -InstanceId i-1234567890abcdef0 -Affinity host -HostId h-012a3456b7890cdef
  ```

### Viewing Dedicated Hosts

You can view details about a Dedicated Host and the individual instances on it.

#### To view details of instances on a Dedicated Host using the Amazon EC2 console

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Dedicated Hosts**.
3. On the **Dedicated Hosts** page, select the host to view more information about.
4. For information about the host, choose **Description**. For information about instances running on the host, choose **Instances**.

#### To view details of instances on a Dedicated Host using the command line tools

Use one of the following commands:

- **describe-hosts** (AWS CLI)
  
  ```bash
  aws ec2 describe-hosts --host-id host_id
  ```

- **Get-EC2Host** (AWS Tools for Windows PowerShell)
  
  ```powershell
  PS C:\> Get-EC2Host -HostId host_id
  ```

### Tagging Dedicated Hosts

You can assign custom tags to your existing Dedicated Hosts to categorize them in different ways, for example, by purpose, owner, or environment. This helps you to quickly find a specific Dedicated Host based on the custom tags that you assigned. Dedicated Host tags can also be used for cost allocation tracking.

You can also apply tags to Dedicated Hosts at the time of creation. For more information, see [Allocating Dedicated Hosts](p. 298).

You can tag a Dedicated Host using the Amazon EC2 console and command line tools.
To tag a Dedicated Host using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Dedicated Hosts.
3. Select the Dedicated Host to tag, and then choose Tags.
4. Choose Add/Edit Tags.
5. In the Add/Edit Tags dialog box, choose Create Tag, and then specify the key and value for the tag.
6. (Optional) Choose Create Tag to add additional tags to the Dedicated Host.
7. Choose Save.

To tag a Dedicated Host using the command line

Use one of the following commands:

- create-tags (AWS CLI)
  
The following command tags the specified Dedicated Host with Owner=TeamA.
  
  ```
  aws ec2 create-tags --resources h-abc12345678909876 --tags Key=Owner,Value=TeamA
  ```

- New-EC2Tag (AWS Tools for Windows PowerShell)
  
The New-EC2Tag command needs a Tag object, which specifies the key and value pair to be used for the Dedicated Host tag. The following commands create a Tag object named $tag, with a key and value pair of Owner and TeamA respectively.
  
  ```
  PS C:\> $tag = New-Object Amazon.EC2.Model.Tag
  PS C:\> $tag.Key = "Owner"
  PS C:\> $tag.Value = "TeamA"
  ```
  
The following command tags the specified Dedicated Host with the $tag object.
  
  ```
  PS C:\> New-EC2Tag -Resource h-abc12345678909876 -Tag $tag
  ```

Monitoring Dedicated Hosts

Amazon EC2 constantly monitors the state of your Dedicated Hosts. Updates are communicated on the Amazon EC2 console. You can also obtain information about your Dedicated Hosts by using the command line tools.

To view the state of a Dedicated Host using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Dedicated Hosts.
3. Locate the Dedicated Host in the list and review the value in the State column.

To view the state of a Dedicated Host using the command line tools

Use one of the following commands and then review the state property in the hostSet response element:

- describe-hosts (AWS CLI)
The following table explains the possible Dedicated Host states.

<table>
<thead>
<tr>
<th>State</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>available</td>
<td>AWS hasn’t detected an issue with the Dedicated Host. No maintenance or repairs are scheduled. Instances can be launched onto this Dedicated Host.</td>
</tr>
<tr>
<td>released</td>
<td>The Dedicated Host has been released. The host ID is no longer in use. Released hosts cannot be reused.</td>
</tr>
<tr>
<td>under-assessment</td>
<td>AWS is exploring a possible issue with the Dedicated Host. If action must be taken, you are notified via the AWS Management Console or email. Instances cannot be launched onto a Dedicated Host in this state.</td>
</tr>
<tr>
<td>pending</td>
<td>A host recovery is in progress. For more information, see Host Recovery (p. 307).</td>
</tr>
<tr>
<td>permanent-failure</td>
<td>An unrecoverable failure has been detected. You receive an eviction notice through your instances and by email. Your instances might continue to run. If you stop or terminate all instances on a Dedicated Host with this state, AWS retires the host. AWS does not restart instances in this state. Instances cannot be launched onto Dedicated Hosts in this state.</td>
</tr>
<tr>
<td>released-permanent-failure</td>
<td>AWS permanently releases Dedicated Hosts that have failed and no longer have running instances on them. The Dedicated Host ID is no longer available for use.</td>
</tr>
</tbody>
</table>

**Releasing Dedicated Hosts**

Any running instances on the Dedicated Host must be stopped before you can release the host. These instances can be migrated to other Dedicated Hosts in your account so that you can continue to use them. These steps apply only to On-Demand Dedicated Hosts.

**To release a Dedicated Host using the Amazon EC2 console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Dedicated Hosts in the navigation pane.
3. On the Dedicated Hosts page, select the Dedicated Host to release.
5. Choose Release to confirm.

**To release a Dedicated Host using the command line tools**

Use one of the following commands:

- release-hosts (AWS CLI)
After you release a Dedicated Host, you cannot reuse the same host or host ID again, and you are no longer charged On-Demand billing rates for it. The Dedicated Host's state is changed to `released`, and you are not able to launch any instances onto that host.

**Note**
If you have recently released Dedicated Hosts, it may take some time for them to stop counting towards your limit. During this time, you may experience `LimitExceeded` errors when trying to allocate new Dedicated Hosts. If this is the case, try allocating new hosts again after a few minutes.

The instances that were stopped are still available for use and are listed on the `Instances` page. They retain their host tenancy setting.

**Purchasing Dedicated Host Reservations**

You can purchase reservations using the Amazon EC2 console or command line tools.

**To purchase reservations using the Amazon EC2 console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose `Dedicated Hosts`, `Dedicated Host Reservations`, `Purchase Dedicated Host Reservation`.
3. On the `Purchase Dedicated Host Reservation` screen, you can search for available offerings using the default settings, or you can specify custom values for the following:
   - **Host instance family**—The options listed correspond with the Dedicated Hosts in your account that are not already assigned to a reservation.
   - **Availability Zone**—The Availability Zone of the Dedicated Hosts in your account that aren't already assigned to a reservation.
   - **Payment option**—The payment option for the offering.
   - **Term**—The term of the reservation, which can be one or three years.
4. Choose `Find offering` and select an offering that matches your requirements.
5. Choose the Dedicated Hosts to associate with the reservation, and then choose `Review`.
6. Review your order and choose `Order`.

**To purchase reservations using the command line tools**

1. Use one of the following commands to list the available offerings that match your needs. The following examples list the offerings that support instances in the `m4` instance family and have a one-year term.
   **Note**
The term is specified in seconds. A one-year term includes 31536000 seconds, and a three-year term includes 94608000 seconds.
   - **describe-host-reservation-offerings** (AWS CLI)

```bash
aws ec2 describe-host-reservation-offerings --filter Name=instance-family,Values=m4 --max-duration 31536000
```
• **Get-EC2HostReservationOffering** (AWS Tools for Windows PowerShell)

```powershell
PS C:\> #filter = @{Name="instance-family"; Value="m4"}

PS C:\> Get-EC2HostReservationOffering -filter $filter -MaxDuration 31536000
```

Both commands return a list of offerings that match your criteria. Note the offeringId of the offering to purchase.

2. Use one of the following commands to purchase the offering and provide the offeringId noted in the previous step. The following examples purchase the specified reservation and associate it with a specific Dedicated Host that is already allocated in the AWS account.

• **purchase-host-reservation** (AWS CLI)

```bash
aws ec2 purchase-host-reservation --offering-id hro-03f707bf363b6b324 --host-idset h-013abcd2a00c0bd123
```

• **New-EC2HostReservation** (AWS Tools for Windows PowerShell)

```powershell
PS C:\> New-EC2HostReservation -OfferingId hro-03f707bf363b6b324 -HostIdSet h-013abcd2a00c0bd123
```

**Viewing Dedicated Host Reservations**

You can view information about the Dedicated Hosts associated with your reservation, including:

- The term of the reservation
- The payment option
- The start and end dates

**To view details of reservations using the Amazon EC2 console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose **Dedicated Hosts** in the navigation pane.
3. On the **Dedicated Hosts** page, choose **Dedicated Host Reservations**, and then select the reservation from the list provided.
4. Choose **Details** for information about the reservation.
5. Choose **Hosts** for information about the Dedicated Hosts with which the reservation is associated.

**To view details of reservations using the command line tools**

Use one of the following commands:

• **describe-host-reservations** (AWS CLI)

```bash
aws ec2 describe-host-reservations
```

• **Get-EC2HostReservation** (AWS Tools for Windows PowerShell)

```powershell
PS C:\> Get-EC2HostReservation
```
Tagging Dedicated Host Reservations

You can assign custom tags to your Dedicated Host Reservations to categorize them in different ways, for example, by purpose, owner, or environment. This helps you to quickly find a specific Dedicated Host Reservation based on the custom tags you assigned.

You can tag a Dedicated Host Reservation using the AWS CLI only.

To tag a Dedicated Host Reservation using the command line

Use one of the following commands:

- create-tags (AWS CLI)

  aws ec2 create-tags --resources hr-1234563a4ff669ae --tags Key=Owner,Value=TeamA

- New-EC2Tag (AWS Tools for Windows PowerShell)

  The New-EC2Tag command needs a Tag parameter, which specifies the key and value pair to be used for the Dedicated Host Reservation tag. The following commands create the Tag parameter.

  PS C:\> $tag = New-Object Amazon.EC2.Model.Tag
  PS C:\> $tag.Key = "Owner"
  PS C:\> $tag.Value = "TeamA"

  PS C:\> New-EC2Tag -Resource hr-1234563a4ff669ae -Tag $tag

Host Recovery

Host recovery automatically restarts your instances on to a new replacement host if failures are detected on your Dedicated Host. Host recovery reduces the need for manual intervention and lowers the operational burden if there is an unexpected Dedicated Host failure.

Additionally, built-in integration with AWS License Manager automates the tracking and management of your licenses if a host recovery occurs.

Note

AWS License Manager integration is supported only in Regions in which AWS License Manager is available.

Contents

- Host Recovery Basics (p. 307)
- Configuring Host Recovery (p. 308)
- Host Recovery States (p. 310)
- Supported Instance Configurations (p. 310)
- Manually Recovering Unsupported Instances (p. 310)
- Related Services (p. 311)
- Pricing (p. 311)

Host Recovery Basics

Host recovery uses host-level health checks to assess Dedicated Host availability and to detect underlying system failures. Examples of problems that can cause host-level health checks to fail include:
• Loss of network connectivity
• Loss of system power
• Hardware or software issues on the physical host

When a system failure is detected on your Dedicated Host, host recovery is initiated and Amazon EC2 automatically allocates a replacement Dedicated Host. The replacement Dedicated Host receives a new host ID, but retains the same attributes as the original Dedicated Host, including:

• Availability Zone
• Instance type
• Tags
• Auto placement settings

After the replacement Dedicated Host is allocated, the instances are recovered on to the replacement Dedicated Host. The recovered instances retain the same attributes as the original instances, including:

• Instance ID
• Private IP addresses
• Elastic IP addresses
• EBS volume attachments
• All instance metadata

If instances have a host affinity relationship with the impaired Dedicated Host, the recovered instances establish host affinity with the replacement Dedicated Host.

When all of the instances have been recovered on to the replacement Dedicated Host, the impaired Dedicated Host is released, and the replacement Dedicated Host becomes available for use.

When host recovery is initiated, the AWS account owner is notified by email and by an AWS Personal Health Dashboard event. A second notification is sent after the host recovery has been successfully completed.

Stopped instances are not recovered on to the replacement Dedicated Host. If you attempt to start a stopped instance that targets the impaired Dedicated Host, the instance start fails. We recommend that you modify the stopped instance to either target a different Dedicated Host, or to launch on any available Dedicated Host with matching configurations and auto-placement enabled.

Instances with instance storage are not recovered on to the replacement Dedicated Host. As a remedial measure, the impaired Dedicated Host is marked for retirement and you receive a retirement notification after the host recovery is complete. Follow the remedial steps described in the retirement notification within the specified time period to manually recover the remaining instances on the impaired Dedicated Host.

If you are using AWS License Manager to track your licenses, AWS License Manager allocates new licenses for the replacement Dedicated Host based on the license configuration limits. If the license configuration has hard limits that will be breached as a result of the host recovery, the recovery process is not allowed and you are notified of the host recovery failure through an Amazon SNS notification. If the license configuration has soft limits that will be breached as a result of the host recovery, the recovery is allowed to continue and you are notified of the limit breach through an Amazon SNS notification. For more information, see Using License Configurations in the AWS License Manager User Guide.

Configuring Host Recovery

You can configure host recovery at the time of Dedicated Host allocation, or after allocation using the Amazon EC2 console or AWS Command Line Interface (CLI).
Enabling Host Recovery

You can enable host recovery at the time of Dedicated Host allocation or after allocation.

For more information about enabling host recovery at the time of Dedicated Host allocation, see Allocating Dedicated Hosts (p. 298).

To enable host recovery after allocation (Console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Dedicated Hosts.
3. Select the Dedicated Host for which to enable host recovery, and then choose Actions, Modify Host Recovery.
4. For Host recovery, choose Enable, and then choose Save.

To enable host recovery after allocation (AWS CLI)

Use the modify-hosts command and specify the host-recovery parameter.

```
$ aws ec2 modify-hosts --host-recovery on --host-ids h-1234567890
```

Disabling Host Recovery

You can disable host recovery at any time after the Dedicated Host has been allocated.

To disable host recovery after allocation (Console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Dedicated Hosts.
3. Select the Dedicated Host for which to disable host recovery, and then choose Actions, Modify Host Recovery.
4. For Host recovery, choose Disable, and then choose Save.

To disable host recovery after allocation (AWS CLI)

Use the modify-hosts command and specify the host-recovery parameter.

```
$ aws ec2 modify-hosts --host-recovery off --host-ids h-1234567890
```

Viewing Host Recovery Configuration

You can view the host recovery configuration for a Dedicated Host at any time.

To view the host recovery configuration for a Dedicated Host (Console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Dedicated Hosts.
3. Select the Dedicated Host, and in the Description tab, review the Host Recovery field.
To view the host recovery configuration for a Dedicated Host (AWS CLI)

Use the `describe-hosts` command.

```bash
$ aws ec2 describe-hosts --host-ids h-1234567890
```

The `HostRecovery` response element indicates whether host recovery is enabled or disabled.

**Host Recovery States**

When a Dedicated Host failure is detected, the impaired Dedicated Host enters the `under-assessment` state, and all of the instances enter the `impaired` state. You cannot launch instances on to the impaired Dedicated Host while it is in the `under-assessment` state.

After the replacement Dedicated Host is allocated, it enters the `pending` state. It remains in this state until the host recovery process is complete. You cannot launch instances on to the replacement Dedicated Host while it is in the `pending` state. Recovered instances on the replacement Dedicated Host remain in the `impaired` state during the recovery process.

After the host recovery is complete, the replacement Dedicated Host enters the `available` state, and the recovered instances return to the `running` state. You can launch instances on to the replacement Dedicated Host after it enters the `available` state. The original impaired Dedicated Host is permanently released and it enters the `released-permanent-failure` state.

If the impaired Dedicated Host has instances that do not support host recovery, such as instances with instance store-backed volumes, the Dedicated Host is not released. Instead, it is marked for retirement and enters the `permanent-failure` state.

**Supported Instance Configurations**

Host recovery is supported only with instances with a supported configuration. To recover instances that are not supported, see Manually Recovering Unsupported Instances (p. 310).

The following instance configurations are not supported:

- Instance store volumes
  - C5d, G4, I3en, M5ad, M5d, P3dn, R5ad, R5d, and z1d instances
  - D2, F1, HS1, I2, I3, X1, and X1e instances
- Bare metal instances

**Manually Recovering Unsupported Instances**

Host recovery does not support recovering instances that use instance store volumes. Follow the instructions below to manually recover any of your instances that could not be automatically recovered.

**Warning**

Data on instance store volumes is lost when an instance is stopped or terminated. This includes instance store volumes that are attached to an instance that has an EBS volume as the root device. To keep data from instance store volumes, back it up to persistent storage before the instance is stopped or terminated.

**Manually Recovering EBS-Backed Instances**

For EBS-backed instances that could not be automatically recovered, we recommend that you manually stop and start the instances to recover them onto a new Dedicated Host. For more information about stopping your instance, and about the changes that occur in your instance configuration when it's stopped, see Stop and Start Your Instance (p. 383).
Manually Recovering Instance Store-Backed Instances

For instance store-backed instances that could not be automatically recovered, we recommend that you do the following:

1. Launch a replacement instance on a new Dedicated Host from your most recent AMI.
2. Migrate all of the necessary data to the replacement instance.
3. Terminate the original instance on the impaired Dedicated Host.

Related Services

Dedicated Host integrates with the following AWS services:

- **AWS License Manager**—Tracks licenses across your Amazon EC2 Dedicated Hosts (supported only in Regions in which AWS License Manager is available). For more information, see the [AWS License Manager User Guide](#).

Pricing

There are no additional charges for using host recovery, but the usual Dedicated Host charges apply. For more information, see [Amazon EC2 Dedicated Hosts Pricing](#).

As soon as host recovery is initiated, you are no longer billed for the impaired Dedicated Host. Billing for the replacement Dedicated Host begins only after it enters the available state.

If the impaired Dedicated Host was billed using the On-Demand rate, the replacement Dedicated Host is also billed using the On-Demand rate. If the impaired Dedicated Host had an active Dedicated Host Reservation, it is transferred to the replacement Dedicated Host.

Tracking Configuration Changes

You can use AWS Config to record configuration changes for Dedicated Hosts, and instances that are launched, stopped, or terminated on them. You can then use the information captured by AWS Config as a data source for license reporting.

AWS Config records configuration information for Dedicated Hosts and instances individually and pairs this information through relationships. There are three reporting conditions:

- **AWS Config recording status**—When On, AWS Config is recording one or more AWS resource types, which can include Dedicated Hosts and Dedicated Instances. To capture the information required for license reporting, verify that hosts and instances are being recorded with the following fields.

- **Host recording status**—When Enabled, the configuration information for Dedicated Hosts is recorded.

- **Instance recording status**—When Enabled, the configuration information for Dedicated Instances is recorded.

If any of these three conditions are disabled, the icon in the **Edit Config Recording** button is red. To derive the full benefit of this tool, ensure that all three recording methods are enabled. When all three are enabled, the icon is green. To edit the settings, choose **Edit Config Recording**. You are directed to the **Set up AWS Config** page in the AWS Config console, where you can set up AWS Config and start recording for your hosts, instances, and other supported resource types. For more information, see [Setting up AWS Config using the Console](#) in the [AWS Config Developer Guide](#).

**Note**

AWS Config records your resources after it discovers them, which might take several minutes.

After AWS Config starts recording configuration changes to your hosts and instances, you can get the configuration history of any host that you have allocated or released and any instance that you have
launched, stopped, or terminated. For example, at any point in the configuration history of a Dedicated Host, you can look up how many instances are launched on that host, along with the number of sockets and cores on the host. For any of those instances, you can also look up the ID of its Amazon Machine image (AMI). You can use this information to report on licensing for your own server-bound software that is licensed per-socket or per-core.

You can view configuration histories in any of the following ways.

- By using the AWS Config console. For each recorded resource, you can view a timeline page, which provides a history of configuration details. To view this page, choose the gray icon in the **Config Timeline** column of the **Dedicated Hosts** page. For more information, see **Viewing Configuration Details in the AWS Config Console** in the **AWS Config Developer Guide**.

- By running AWS CLI commands. First, you can use the `list-discovered-resources` command to get a list of all hosts and instances. Then, you can use the `get-resource-config-history` command to get the configuration details of a host or instance for a specific time interval. For more information, see **View Configuration Details Using the CLI** in the **AWS Config Developer Guide**.

- By using the AWS Config API in your applications. First, you can use the `ListDiscoveredResources` action to get a list of all hosts and instances. Then, you can use the `GetResourceConfigHistory` action to get the configuration details of a host or instance for a specific time interval.

For example, to get a list of all of your Dedicated Hosts from AWS Config, run a CLI command such as the following.

```bash
aws configservice list-discovered-resources --resource-type AWS::EC2::Host
```

To obtain the configuration history of a Dedicated Host from AWS Config, run a CLI command such as the following.

```bash
aws configservice get-resource-config-history --resource-type AWS::EC2::Instance --resource-id i-1234567890abcdef0
```

**To manage AWS Config settings using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the **Dedicated Hosts** page, choose **Edit Config Recording**.
3. In the AWS Config console, follow the steps provided to turn on recording. For more information, see **Setting up AWS Config using the Console**.

For more information, see **Viewing Configuration Details in the AWS Config Console**.

**To activate AWS Config using the command line or API**

- Using the AWS CLI, see **Viewing Configuration Details (AWS CLI)** in the **AWS Config Developer Guide**.
- Using the Amazon EC2 API, see **GetResourceConfigHistory**.

**Dedicated Instances**

Dedicated Instances are Amazon EC2 instances that run in a virtual private cloud (VPC) on hardware that's dedicated to a single customer. Dedicated Instances that belong to different AWS accounts are physically isolated at the hardware level. In addition, Dedicated Instances that belong to AWS accounts that are linked to a single payer account are also physically isolated at the hardware level. However, Dedicated Instances may share hardware with other instances from the same AWS account that are not Dedicated Instances.
Note

A Dedicated Host is also a physical server that's dedicated for your use. With a Dedicated Host, you have visibility and control over how instances are placed on the server. For more information, see Dedicated Hosts (p. 294).

Dedicated Instance Basics

Each instance that you launch into a VPC has a tenancy attribute. This attribute has the following values.

<table>
<thead>
<tr>
<th>Tenancy Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td>Your instance runs on shared hardware.</td>
</tr>
<tr>
<td>dedicated</td>
<td>Your instance runs on single-tenant hardware.</td>
</tr>
<tr>
<td>host</td>
<td>Your instance runs on a Dedicated Host, which is an isolated server with configurations that you can control.</td>
</tr>
</tbody>
</table>

After you launch an instance, there are some limitations to changing its tenancy.

- You cannot change the tenancy of an instance from default to dedicated or host after you've launched it.
- You cannot change the tenancy of an instance from dedicated or host to default after you've launched it.

You can change the tenancy of an instance from dedicated to host, or from host to dedicated after you've launched it. For more information, see Changing the Tenancy of an Instance (p. 316).

Each VPC has a related instance tenancy attribute. This attribute has the following values.

<table>
<thead>
<tr>
<th>Tenancy Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td>An instance launched into the VPC runs on shared hardware by default, unless you explicitly specify a different tenancy during instance launch.</td>
</tr>
<tr>
<td>dedicated</td>
<td>An instance launched into the VPC is a Dedicated Instance by default, unless you explicitly specify a tenancy of host during instance launch. You cannot specify a tenancy of default during instance launch.</td>
</tr>
</tbody>
</table>

You can change the instance tenancy of a VPC from dedicated to default after you create it. You cannot change the instance tenancy of a VPC to dedicated.

To create Dedicated Instances, you can do the following:

- Create the VPC with the instance tenancy set to dedicated (all instances launched into this VPC are Dedicated Instances).
- Create the VPC with the instance tenancy set to default, and specify a tenancy of dedicated for any instances when you launch them.

Dedicated Instances Limitations

Some AWS services or their features won't work with a VPC with the instance tenancy set to dedicated. Check the service's documentation to confirm if there are any limitations.
Some instance types cannot be launched into a VPC with the instance tenancy set to dedicated. For more information about supported instances types, see Amazon EC2 Dedicated Instances.

**Amazon EBS with Dedicated Instances**

When you launch an Amazon EBS-backed Dedicated Instance, the EBS volume doesn't run on single-tenant hardware.

**Reserved Instances with Dedicated Tenancy**

To guarantee that sufficient capacity is available to launch Dedicated Instances, you can purchase Dedicated Reserved Instances. For more information, see Reserved Instances (p. 197).

When you purchase a Dedicated Reserved Instance, you are purchasing the capacity to launch a Dedicated Instance into a VPC at a much reduced usage fee; the price break in the usage charge applies only if you launch an instance with dedicated tenancy. When you purchase a Reserved Instance with default tenancy, it applies only to a running instance with default tenancy; it would not apply to a running instance with dedicated tenancy.

You can't use the modification process to change the tenancy of a Reserved Instance after you've purchased it. However, you can exchange a Convertible Reserved Instance for a new Convertible Reserved Instance with a different tenancy.

**Automatic Scaling of Dedicated Instances**

You can use Amazon EC2 Auto Scaling to launch Dedicated Instances. For more information, see Launching Auto Scaling Instances in a VPC in the Amazon EC2 Auto Scaling User Guide.

**Automatic Recovery of Dedicated Instances**

You can configure automatic recovery for a Dedicated Instances if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair. For more information, see Recover Your Instance (p. 393).

**Dedicated Spot Instances**

You can run a Dedicated Spot Instance by specifying a tenancy of dedicated when you create a Spot Instance request. For more information, see Specifying a Tenancy for Your Spot Instances (p. 246).

**Pricing for Dedicated Instances**

Pricing for Dedicated Instances is different to pricing for On-Demand Instances. For more information, see the Amazon EC2 Dedicated Instances product page.

**Working with Dedicated Instances**

You can create a VPC with an instance tenancy of dedicated to ensure that all instances launched into the VPC are Dedicated Instances. Alternatively, you can specify the tenancy of the instance during launch.

**Topics**

- Creating a VPC with an Instance Tenancy of Dedicated (p. 315)
- Launching Dedicated Instances into a VPC (p. 315)
- Displaying Tenancy Information (p. 316)
- Changing the Tenancy of an Instance (p. 316)
- Changing the Tenancy of a VPC (p. 317)
Creating a VPC with an Instance Tenancy of Dedicated

When you create a VPC, you have the option of specifying its instance tenancy. If you're using the Amazon VPC console, you can create a VPC using the VPC wizard or the Your VPCs page.

To create a VPC with an instance tenancy of dedicated (VPC Wizard)
1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. From the dashboard, choose Start VPC Wizard.
3. Select a VPC configuration, and then choose Select.
4. On the next page of the wizard, choose Dedicated from the Hardware tenancy list.
5. Choose Create VPC.

To create a VPC with an instance tenancy of dedicated (Create VPC dialog box)
1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose Your VPCs, and then Create VPC.
3. For Tenancy, choose Dedicated. Specify the CIDR block, and choose Yes, Create.

To set the tenancy option when you create a VPC using the command line
- create-vpc (AWS CLI)
- New-EC2Vpc (AWS Tools for Windows PowerShell)

If you launch an instance into a VPC that has an instance tenancy of dedicated, your instance is automatically a Dedicated Instance, regardless of the tenancy of the instance.

Launching Dedicated Instances into a VPC

You can launch a Dedicated Instance using the Amazon EC2 launch instance wizard.

To launch a Dedicated Instance into a default tenancy VPC using the console
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Launch Instance.
3. On the Choose an Amazon Machine Image (AMI) page, select an AMI and choose Select.
4. On the Choose an Instance Type page, select the instance type and choose Next: Configure Instance Details.
   Note
   Ensure that you choose an instance type that's supported as a Dedicated Instance. For more information, see Amazon EC2 Dedicated Instances.
5. On the Configure Instance Details page, select a VPC and subnet. Choose Dedicated - Run a dedicated instance from the Tenancy list, and then Next: Add Storage.
6. Continue as prompted by the wizard. When you've finished reviewing your options on the Review Instance Launch page, choose Launch to choose a key pair and launch the Dedicated Instance.

For more information about launching an instance with a tenancy of host, see Launching Instances onto Dedicated Hosts (p. 299).

To set the tenancy option for an instance during launch using the command line
- run-instances (AWS CLI)
Displaying Tenancy Information

To display tenancy information for your VPC using the console

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose Your VPCs.
3. Check the instance tenancy of your VPC in the Tenancy column.
4. If the Tenancy column is not displayed, choose Edit Table Columns (the gear-shaped icon), Tenancy in the Show/Hide Columns dialog box, and then Close.

To display tenancy information for your instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Check the tenancy of your instance in the Tenancy column.
4. If the Tenancy column is not displayed, do one of the following:
   - Choose Show/Hide Columns (the gear-shaped icon), Tenancy in the Show/Hide Columns dialog box, and then Close.
   - Select the instance. The Description tab in the details pane displays information about the instance, including its tenancy.

To describe the tenancy of your VPC using the command line

- describe-vpcs (AWS CLI)
- Get-EC2Vpc (AWS Tools for Windows PowerShell)

To describe the tenancy of your instance using the command line

- describe-instances (AWS CLI)
- Get-EC2Instance (AWS Tools for Windows PowerShell)

To describe the tenancy value of a Reserved Instance using the command line

- describe-reserved-instances (AWS CLI)
- Get-EC2ReservedInstance (AWS Tools for Windows PowerShell)

To describe the tenancy value of a Reserved Instance offering using the command line

- describe-reserved-instances-offerings (AWS CLI)
- Get-EC2ReservedInstancesOffering (AWS Tools for Windows PowerShell)

Changing the Tenancy of an Instance

Depending on your instance type and platform, you can change the tenancy of a stopped Dedicated Instance to host after launching it. The next time the instance starts, it's started on a Dedicated Host that's allocated to your account. For more information about allocating and working with Dedicated
Hosts, and the instance types that can be used with Dedicated Hosts, see Working with Dedicated Hosts (p. 297). Similarly, you can change the tenancy of a stopped Dedicated Host instance to dedicated after launching it. The next time the instance starts, it's started on single-tenant hardware that we control.

To change the tenancy of an instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances and select your instance.
3. Choose Actions, Instance State, Stop.
5. In the Tenancy list, choose whether to run your instance on dedicated hardware or on a Dedicated Host. Choose Save.

To modify the tenancy value of an instance using the command line

- modify-instance-placement (AWS CLI)
- Edit-EC2InstancePlacement (AWS Tools for Windows PowerShell)

Changing the Tenancy of a VPC

You can change the instance tenancy attribute of a VPC from dedicated to default. Modifying the instance tenancy of the VPC does not affect the tenancy of any existing instances in the VPC. The next time you launch an instance in the VPC, it has a tenancy of default, unless you specify otherwise during launch.

You cannot change the instance tenancy attribute of a VPC to dedicated.

You can modify the instance tenancy attribute of a VPC using the AWS CLI, an AWS SDK, or the Amazon EC2 API only.

To modify the instance tenancy attribute of a VPC using the AWS CLI

- Use the modify-vpc-tenancy command to specify the ID of the VPC and instance tenancy value. The only supported value is default.

```
aws ec2 modify-vpc-tenancy --vpc-id vpc-1a2b3c4d --instance-tenancy default
```

On-Demand Capacity Reservations

On-Demand Capacity Reservations enable you to reserve capacity for your Amazon EC2 instances in a specific Availability Zone for any duration. This gives you the ability to create and manage capacity reservations independently from the billing discounts offered by Reserved Instances (RI). By creating Capacity Reservations, you ensure that you always have access to EC2 capacity when you need it, for as long as you need it. Capacity Reservations can be created at any time, without entering into a one-year or three-year term commitment, and the capacity is available immediately. When you no longer need the reservation, cancel the Capacity Reservation to stop incurring charges for it.

When you create a Capacity Reservation, you specify the Availability Zone in which you want to reserve the capacity, the number of instances for which you want to reserve capacity, and the instance attributes, including the instance type, tenancy, and platform/OS. Capacity Reservations can only be used by instances that match their attributes. By default, they are automatically used by show running instances
that match the attributes. If you don't have any running instances that match the attributes of the Capacity Reservation, it remains unused until you launch an instance with matching attributes.

In addition, you can use your Regional RIs with your Capacity Reservations to benefit from billing discounts. This gives you the flexibility to selectively add capacity reservations and still get the Regional RI discounts for that usage. AWS automatically applies your RI discount when the attributes of a Capacity Reservation match the attributes of an active Regional RI.

Contents
- Differences between Capacity Reservations and RIs (p. 318)
- Capacity Reservation Limits (p. 318)
- Capacity Reservation Limitations and Restrictions (p. 318)
- Capacity Reservation Pricing and Billing (p. 319)
- Working with Capacity Reservations (p. 320)
- Working with Shared Capacity Reservations (p. 324)

Differences between Capacity Reservations and RIs

The following table highlights some key differences between Capacity Reservations and RIs:

<table>
<thead>
<tr>
<th></th>
<th>Capacity Reservations</th>
<th>Zonal RIs</th>
<th>Regional RIs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Term</strong></td>
<td>No commitment required. Can be created and canceled as needed.</td>
<td>Require fixed one-year or three-year commitment.</td>
<td></td>
</tr>
<tr>
<td><strong>Capacity benefit</strong></td>
<td>Reserve capacity in a specific Availability Zone.</td>
<td>Reserve capacity in a specific Availability Zone.</td>
<td>Do not reserve capacity in an Availability Zone.</td>
</tr>
<tr>
<td><strong>Billing discount</strong></td>
<td>No billing discount. Instances launched into a Capacity Reservation are charged at their standard On-Demand rates. However, Regional RIs can be used with Capacity Reservations to get a billing discount.</td>
<td>Provide billing discounts.</td>
<td></td>
</tr>
<tr>
<td><strong>Instance Limits</strong></td>
<td>Limited to your On-Demand Instance limits per Region.</td>
<td>Limited to 20 per Availability Zone. A limit increase can be requested.</td>
<td>Limited to 20 per Region. A limit increase can be requested.</td>
</tr>
</tbody>
</table>

Capacity Reservation Limits

The number of instances for which you are allowed to reserve capacity is based on your account's On-Demand Instance limit. You can reserve capacity for as many instances as that limit allows, minus the number of instances that are already running.

Capacity Reservation Limitations and Restrictions

Before you create Capacity Reservations, take note of the following limitations and restrictions.

- Active and unused Capacity Reservations count towards your On-Demand Instance limits
• Capacity Reservations are not transferable from one AWS account to another
• Zonal RI billing discounts do not apply to Capacity Reservations
• Capacity Reservations can't be created in placement groups
• Capacity Reservations can't be used with Dedicated Hosts

Capacity Reservation Pricing and Billing

Pricing

When the Capacity Reservation is active, you are charged the equivalent On-Demand rate whether you run the instances or not. If you do not use the reservation, this shows up as unused reservation on your EC2 bill. When you run an instance that matches the attributes of a reservation, you just pay for the instance and nothing for the reservation. There are no upfront or additional charges.

For example, if you create a Capacity Reservation for 20 m4.large Linux instances and run 15 m4.large Linux instances in the same Availability Zone, you will be charged for 15 active instances and for 5 unused instances in the reservation.

Regional RIs billing discounts apply to Capacity Reservations. AWS automatically applies your active Regional RIs to active and unused Capacity Reservations that have matching attributes. For more information about Regional RIs, see Reserved Instances (p. 197).

For more information about Amazon EC2 pricing, see Amazon EC2 Pricing.

Billing

Capacity Reservations are billed at per-second granularity. This means that you are charged for partial hours. For example, if a reservation remains active in your account for 24 hours and 15 minutes, you will be billed for 24.25 reservation hours.

The following example shows how a Capacity Reservation is billed. The Capacity Reservation is created for one m4.large Linux instance, which has an On-Demand rate of $0.10 per usage hour. In this example, the Capacity Reservation is active in the account for five hours. The Capacity Reservation is unused for the first hour, so it is billed for one unused hour at the m4.large instance type's standard On-Demand rate. In hours two through five, the Capacity Reservation is occupied by an m4.large instance. During this time, the Capacity Reservation accrues no charges, and the account is instead billed for the m4.large instance occupying it. In the sixth hour, the Capacity Reservation is canceled and the m4.large instance runs normally outside of the reserved capacity. For that hour, it is charged at the On-Demand rate of the m4.large instance type.

<table>
<thead>
<tr>
<th>Hour</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unused Capacity Reservation</td>
<td>$0.10</td>
<td>$0.00</td>
<td>$0.00</td>
<td>$0.00</td>
<td>$0.00</td>
</tr>
<tr>
<td>On-demand Instance Usage</td>
<td>$0.00</td>
<td>$0.10</td>
<td>$0.10</td>
<td>$0.10</td>
<td>$0.10</td>
</tr>
<tr>
<td>Hourly cost</td>
<td>$0.10</td>
<td>$0.10</td>
<td>$0.10</td>
<td>$0.10</td>
<td>$0.10</td>
</tr>
</tbody>
</table>

Billing Discounts

Regional RIs billing discounts apply to Capacity Reservations. AWS automatically applies your active Regional RIs to active Capacity Reservations that have matching attributes. For more information about Regional RIs, see Reserved Instances (p. 197).

Zonal RI billing discounts do not apply to Capacity Reservations.
When your instance-hours and reservation-hours combined exceed your total eligible discounted Regional RI hours, discounts are preferentially applied to instance-hours first and then to unused reservation-hours.

Viewing Your Bill

You can find out about the charges and fees to your account by viewing the AWS Billing and Cost Management console.

- The Dashboard displays a spend summary for your account.
- On the Bills page, under Details, expand the Elastic Compute Cloud section and the Region to get billing information about your Capacity Reservations.

You can view the charges online, or you can download a CSV file. For more information, see Capacity Reservation Line Items in the AWS Billing and Cost Management User Guide.

Working with Capacity Reservations

To start using Capacity Reservations, you need to create the capacity reservation in the required Availability Zone. After you created a Capacity Reservation, you can launch instances into the reserved capacity, view its capacity utilization in real time, and increase or decrease its capacity as needed.

By default, Capacity Reservations automatically match new instances and running instances that have matching attributes (instance type, platform, and Availability Zone). In other words, instances that have matching attributes automatically run in the Capacity Reservation's capacity. However, you can also target a Capacity Reservation for specific workloads. This enables you to explicitly control which instances are allowed to run in that reserved capacity.

You can specify how the reservation ends. You can choose to manually cancel the Capacity Reservation or end it automatically at a specified time. If you specify an end time, the Capacity Reservation is canceled within an hour of the specified time. For example, if you specify 5/31/2019, 13:30:55, the Capacity Reservation is guaranteed to end between 13:30:55 and 14:30:55 on 5/31/2019. After a reservation ends, you can no longer target instances to the Capacity Reservation. Instances running in the reserved capacity continue to run uninterrupted. If instances targeting a Capacity Reservation are stopped, you cannot restart them until you remove their Capacity Reservation targeting preference or configure them to target a different Capacity Reservation.

Contents

- Creating a Capacity Reservation (p. 320)
- Launching Instances into an Existing Capacity Reservation (p. 321)
- Modifying a Capacity Reservation (p. 322)
- Modifying an Instance's Capacity Reservation Settings (p. 323)
- Viewing a Capacity Reservation (p. 323)
- Canceling a Capacity Reservation (p. 324)

Creating a Capacity Reservation

After you create the Capacity Reservation, the capacity is available immediately. The capacity remains reserved for your use as long as the Capacity Reservation is active, and you can launch instances into it at any time. If the Capacity Reservation is open, new instances and existing instances that have matching attributes automatically run in the Capacity Reservation's capacity. If the Capacity Reservation is targeted, instances must specifically target it to run in the reserved capacity.

Your request to create a Capacity Reservation could fail if one of the following is true:
• Amazon EC2 does not have sufficient capacity to fulfill the request. Either try again at a later time, try a different Availability Zone, or try a smaller capacity. If your application is flexible across instance types and sizes, try different instance attributes.

• The requested quantity exceeds your On-Demand Instance limit for the selected instance type. Increase your On-Demand Instance limit for the required instance type and try again. For more information, see Amazon EC2 Service Limits (p. 974).

To create a Capacity Reservation using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Capacity Reservations, choose Create Capacity Reservation.
3. On the Create a Capacity Reservation page, configure the following settings in the Instance details section:
   a. Instance Type—The type of instance to launch into the reserved capacity.
   b. Launch EBS-optimized instances—Specify whether to reserve the capacity for EBS-optimized instances. This option is selected by default for some instance types. For more information about EBS-optimized instances, see Amazon Elastic Block Store (p. 782).
   c. Attach instance store at launch—Specify whether instances launched into the Capacity Reservation use temporary block-level storage. The data on an instance store volume persists only during the life of the associated instance.
   d. Platform—The operating system for your instances.
   e. Availability Zone—The Availability Zone in which to reserve the capacity.
   f. Tenancy—Specify whether to run on shared hardware (default) or a dedicated instance.
   g. Quantity—The number instances for which to reserve capacity. If you specify a quantity that exceeds your remaining On-Demand Instance limit for the selected instance type, the request is denied.
4. Configure the following settings in the Reservation details section:
   a. Reservation Ends—Choose one of the following options:
      • Manually—Reserve the capacity until you explicitly cancel it.
      • Specific time—Cancels the capacity reservation automatically at the specified date and time.
   b. Instance eligibility—Choose one of the following options:
      • open—(Default) The Capacity Reservation matches any instance that has matching attributes (instance type, platform, and Availability Zone). If you launch an instance with matching attributes, it is placed into the reserved capacity automatically.
      • targeted—The Capacity Reservation only accepts instances that have matching attributes (instance type, platform, and Availability Zone), and explicitly target the reservation.
5. Choose Request reservation.

To create a Capacity Reservation using the AWS CLI

Use the create-capacity-reservation command:

```
aws ec2 create-capacity-reservation --instance-type instance_type --instance-platform platform_type --availability-zone az --instance-count quantity
```

Launching Instances into an Existing Capacity Reservation

You can launch instances into an existing Capacity Reservation if it has matching attributes (instance type, platform, and Availability Zone) and sufficient capacity. Launching an instance into a Capacity
Reservation reduces its available capacity by the number of instances launched. For example, if you launch three instances, the Capacity Reservation's available capacity is reduced by three.

**To launch instances into an existing Capacity Reservation using the console**

1. Open the Launch Instance wizard by choosing Launch Instances from Dashboard or Instances.
2. Select an Amazon Machine Image (AMI) and an instance type.
3. Complete the **Configure Instance Details** page. For **Capacity Reservation**, choose one of the following options:
   - **Open** — Launches the instances into any Capacity Reservation that has matching attributes and sufficient capacity for the number of instances you selected. If there is no matching Capacity Reservation with sufficient capacity, the instance uses On-Demand capacity.
   - A specific Capacity Reservation — Launches the instances into this Capacity Reservation. If this Capacity Reservation does not have sufficient capacity for the number of instances you selected, the instance launch fails.
   - **None** — Prevents the instances from launching into a Capacity Reservation.
4. Complete the remaining steps to launch the instances.

**To launch an instance into an existing Capacity Reservation using the AWS CLI**

Use the `run-instances` command and specify the `--capacity-reservation-specification` parameter.

The following example launches a `t2.micro` instance into any open Capacity Reservation that has matching attributes and available capacity:

```bash
aws ec2 run-instances --image-id ami-abc12345 --count 1 --instance-type t2.micro --key-name MyKeyPair --availability-zone us-east-1b --capacity-reservation-specification CapacityReservationPreference=open
```

The following example launches a `t2.micro` instance into a targeted Capacity Reservation:

```bash
aws ec2 run-instances --image-id ami-abc12345 --count 1 --instance-type t2.micro --key-name MyKeyPair --availability-zone us-east-1b --capacity-reservation-specification CapacityReservationTarget= [{CapacityReservationId=cr-a1234567}]
```

**Modifying a Capacity Reservation**

You can change an active Capacity Reservation's attributes after you have created it. You cannot modify a Capacity Reservation after it has expired or after you have explicitly canceled it.

When modifying a Capacity Reservation, you can only increase or decrease the quantity and change the way in which it is released. You cannot change a Capacity Reservation's instance type, EBS optimization, instance store settings, platform, Availability Zone, or instance eligibility. If you need to modify any of these attributes, we recommend that you cancel the reservation, and then create a new one with the required attributes.

If you specify a new quantity that exceeds your remaining On-Demand Instance limit for the selected instance type, the update fails.

**To modify a Capacity Reservation using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. Choose **Capacity Reservations**, select the Capacity Reservation to modify, and then choose **Edit**.
3. Modify the **Quantity** or **Reservation ends** options as needed, and choose **Save changes**.

**To modify a Capacity Reservation using the AWS CLI**

Use the `modify-capacity-reservations` command:

```bash
aws ec2 modify-capacity-reservation --capacity-reservation-id reservation_id --instance-count quantity --end-date-type limited/unlimited --end-date expiration_date
```

**Modifying an Instance's Capacity Reservation Settings**

You can modify the Capacity Reservation settings for a stopped instance at any time as follows:

- Start in any Capacity Reservation that has matching attributes (instance type, platform, and Availability Zone) and available capacity.
- Start the instance in a specific Capacity Reservation.
- Prevent the instance from starting in a Capacity Reservation.

**To modify an instance's Capacity Reservation settings using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. Choose **Instances** and select the instance to modify. Stop the instance if it is not already stopped.
3. Choose **Actions, Modify Capacity Reservation Settings**.
4. For **Capacity Reservation**, choose one of the following options:
   - **Open** — Starts the instance in any open Capacity Reservation that has matching attributes (instance type, platform, and Availability Zone) and available capacity. If you do not have a matching Capacity Reservation with available capacity, the instance uses On-Demand capacity.
   - A specific Capacity Reservation — Runs the instance in this Capacity Reservation. If the instance attributes (instance type, platform, and Availability Zone) do not match those of the Capacity Reservation, or if the selected Capacity Reservation does not have sufficient capacity, the instance launch fails.
   - **None** — Prevents the instance from running in a Capacity Reservation.

**To modify an instance's Capacity Reservation settings using the AWS CLI**

Use the `modify-instance-capacity-reservation-attributes` command:

```bash
aws ec2 modify-instance-capacity-reservation-attributes --instance-id instance_id --capacity-reservation-specification 'CapacityReservationPreference=none|open'
```

**Viewing a Capacity Reservation**

Capacity Reservations have these possible states:

- **active** — The capacity is available for use.
- **expired** — The Capacity Reservation expired automatically at the date and time specified in your reservation request. The reserved capacity is no longer available for your use.
- **cancelled** — The Capacity Reservation was manually canceled. The reserved capacity is no longer available for your use.
• pending—The Capacity Reservation request was successful but the capacity provisioning is still pending.
• failed—The Capacity Reservation request has failed. A request might fail due to invalid request parameters, capacity constraints, or instance limit constraints. You can view a failed request for 60 minutes.

To view your Capacity Reservations using the console
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Capacity Reservations and select a Capacity Reservation to view.
3. Choose View launched instances for this reservation.

To view your Capacity Reservations using the AWS CLI
Use the describe-capacity-reservations command:

```
aws ec2 describe-capacity-reservations
```

Canceling a Capacity Reservation

You can cancel a Capacity Reservation at any time if you no longer need the reserved capacity. When you cancel a Capacity Reservation, the capacity is released immediately, and it is no longer reserved for your use.

You can cancel empty Capacity Reservations and Capacity Reservations that have running instances. If you cancel a Capacity Reservation that has running instances, the instances continue to run normally outside of the capacity reservation at standard On-Demand Instance rates or at a discounted rate if you have an active matching Regional RI.

After you cancel a Capacity Reservation, instances that target it can no longer launch. Modify these instances so that they either target a different Capacity Reservation, launch into any open Capacity Reservation with matching attributes and sufficient capacity, or avoid launching into a Capacity Reservation. For more information, see Modifying an Instance’s Capacity Reservation Settings (p. 323).

To cancel a Capacity Reservation using the console
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Capacity Reservations and select the Capacity Reservation to cancel.
3. Choose Cancel reservation, Cancel reservation.

To cancel a Capacity Reservation using the AWS CLI

Use the cancel-capacity-reservation command:

```
aws ec2 cancel-capacity-reservation --capacity-reservation-id reservation_id
```

Working with Shared Capacity Reservations

Capacity Reservation sharing enables Capacity Reservation owners to share their reserved capacity with other AWS accounts or within an AWS organization. This enables you to create and manage Capacity Reservations centrally, and share the reserved capacity across multiple AWS accounts or within your AWS organization.
In this model, the AWS account that owns the Capacity Reservation (owner) shares it with other AWS accounts (consumers). Consumers can launch instances into Capacity Reservations that are shared with them in the same way that they would with Capacity Reservations that they own in their own account. The Capacity Reservation owner is responsible for managing the Capacity Reservation and the instances that they launch into it. Owners cannot modify instances that consumers launch into Capacity Reservations that they have shared. Consumers are responsible for managing the instances that they launch into Capacity Reservations shared with them. Consumers cannot view or modify instances owned by other consumers or by the Capacity Reservation owner.

A Capacity Reservation owner can share a Capacity Reservation with:

- Specific AWS accounts inside or outside of its AWS organization
- An organizational unit inside its AWS organization
- Its entire AWS organization

Contents

- Prerequisites for Sharing Capacity Reservations (p. 325)
- Related Services (p. 325)
- Sharing Across Availability Zones (p. 326)
- Sharing a Capacity Reservation (p. 326)
- Unsharing a Shared Capacity Reservation (p. 327)
- Identifying a Shared Capacity Reservation (p. 327)
- Viewing Shared Capacity Reservation Usage (p. 328)
- Shared Capacity Reservation Permissions (p. 328)
- Billing and Metering (p. 328)
- Instance Limits (p. 328)

Prerequisites for Sharing Capacity Reservations

- To share a Capacity Reservation, you must own it in your AWS account. You cannot share a Capacity Reservation that has been shared with you.
- You can only share Capacity Reservations for shared tenancy instances. You cannot share Capacity Reservations for dedicated tenancy instances.
- Capacity Reservation sharing is not available to new AWS accounts or AWS accounts that have a limited billing history. New accounts that are linked to a qualified master (payer) account or are linked through an AWS organization are exempt from this restriction.
- To share a Capacity Reservation with your AWS organization or an organizational unit in your AWS organization, you must enable sharing with AWS Organizations. For more information, see Enable Sharing with AWS Organizations in the AWS RAM User Guide.

Related Services

Capacity Reservation sharing integrates with AWS Resource Access Manager (AWS RAM). AWS RAM is a service that enables you to share your AWS resources with any AWS account or through AWS Organizations. With AWS RAM, you share resources that you own by creating a resource share. A resource share specifies the resources to share, and the consumers with whom to share them. Consumers can be individual AWS accounts, or organizational units or an entire organization from AWS Organizations.

For more information about AWS RAM, see the AWS RAM User Guide.
Sharing Across Availability Zones

To ensure that resources are distributed across the Availability Zones for a Region, we independently map Availability Zones to names for each account. This could lead to Availability Zone naming differences across accounts. For example, the Availability Zone `us-east-1a` for your AWS account might not have the same location as `us-east-1a` for another AWS account.

To identify the location of your Capacity Reservations relative to your accounts, you must use the Availability Zone ID (AZ ID). The AZ ID is a unique and consistent identifier for an Availability Zone across all AWS accounts. For example, `use1-az1` is an AZ ID for the `us-east-1` Region and it is the same location in every AWS account.

**To view the AZ IDs for the Availability Zones in your account**

2. The AZ IDs for the current Region are displayed in the Your AZ ID panel on the right-hand side of the screen.

Sharing a Capacity Reservation

When you share a Capacity Reservation that you own with other AWS accounts, you enable them to launch instances into your reserved capacity. If you share an open Capacity Reservation, keep the following in mind as it could lead to unintended Capacity Reservation usage:

- If consumers have running instances that match the Capacity Reservation's attributes, have the CapacityReservationPreference parameter set to open, and are not yet running in reserved capacity, they automatically use the shared Capacity Reservation.
- If consumers launch instances that have matching attributes (instance type, platform, and Availability Zone) and have the CapacityReservationPreference parameter set to open, they automatically launch into the shared Capacity Reservation.

To share a Capacity Reservation, you must add it to a resource share. A resource share is an AWS RAM resource that lets you share your resources across AWS accounts. A resource share specifies the resources to share, and the consumers with whom they are shared. When you share a Capacity Reservation using the Amazon EC2 console, you add it to an existing resource share. To add the Capacity Reservation to a new resource share, you must create the resource share using the AWS RAM console.

If you are part of an organization in AWS Organizations and sharing within your organization is enabled, consumers in your organization are automatically granted access to the shared Capacity Reservation. Otherwise, consumers receive an invitation to join the resource share and are granted access to the shared Capacity Reservation after accepting the invitation.

You can share a Capacity Reservation that you own using the Amazon EC2 console, AWS RAM console, or the AWS CLI.

**To share a Capacity Reservation that you own using the Amazon EC2 console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Capacity Reservations.
3. Choose the Capacity Reservation to share and choose Actions, Share reservation.
4. Select the resource share to which to add the Capacity Reservation and choose Share Capacity Reservation.

*Note*

It could take a few minutes for consumers to get access to the shared Capacity Reservation.
To share a Capacity Reservation that you own using the AWS RAM console


To share a Capacity Reservation that you own using the AWS CLI

Use the create-resource-share command.

Unsharing a Shared Capacity Reservation

The Capacity Reservation owner can unshare a shared Capacity Reservation at any time. When you unshare a shared Capacity Reservation, the following rules apply:

- Instances owned by consumers that were running in the shared capacity at the time of unsharing continue to run normally outside of the reserved capacity, and the capacity is restored to the Capacity Reservation subject to Amazon EC2 capacity availability.
- Consumers with whom the Capacity Reservation was shared can no longer launch new instances into the reserved capacity.

To unshare a shared Capacity Reservation that you own, you must remove it from the resource share. You can do this using the Amazon EC2 console, AWS RAM console, or the AWS CLI.

To unshare a shared Capacity Reservation that you own using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Capacity Reservations.
3. Choose the Capacity Reservation to unshare and choose the Sharing tab.
4. The Sharing tab lists the resource shares to which the Capacity Reservation has been added. Select the resource share from which to remove the Capacity Reservation and choose Remove from resource share.

To unshare a shared Capacity Reservation that you own using the AWS RAM console

See Updating a Resource Share in the AWS RAM User Guide.

To unshare a shared Capacity Reservation that you own using the AWS CLI

Use the disassociate-resource-share command.

Identifying a Shared Capacity Reservation

Owners and consumers can identify shared Capacity Reservations using the Amazon EC2 console and AWS CLI

To identify a shared Capacity Reservation using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Capacity Reservations. The screen lists Capacity Reservations that you own and Capacity Reservations that are shared with you. The Owner column shows the AWS account ID of the Capacity Reservation owner. (me) next to the AWS account ID indicates that you are the owner.

To identify a shared Capacity Reservation using the AWS CLI
Use the `describe-capacity-reservations` command. The command returns the Capacity Reservations that you own and Capacity Reservations that are shared with you. `OwnerId` shows the AWS account ID of the Capacity Reservation owner.

**Viewing Shared Capacity Reservation Usage**

The owner of a shared Capacity Reservation can view its usage at any time using the Amazon EC2 console and the AWS CLI.

**To view Capacity Reservation usage using the Amazon EC2 console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Capacity Reservations**.
3. Select the Capacity Reservation for which to view the usage and choose the **Usage** tab.

   The **AWS account ID** column shows the account IDs of the consumers currently using the Capacity Reservation. The **Launched instances** column shows the number of instances each consumer currently has running in the reserved capacity.

**To view Capacity Reservation usage using the AWS CLI**

Use the `get-capacity-reservation-usage` command. `AccountId` shows the account ID of the account using the Capacity Reservation. `UsedInstanceCount` shows the number of instances the consumer currently has running in the reserved capacity.

**Shared Capacity Reservation Permissions**

**Permissions for Owners**

Owners are responsible for managing and canceling their shared Capacity Reservations. Owners cannot modify instances running in the shared Capacity Reservation that are owned by other accounts. Owners remain responsible for managing instances that they launch into the shared Capacity Reservation.

**Permissions for Consumers**

Consumers are responsible for managing their instances that are running the shared Capacity Reservation. Consumers cannot modify the shared Capacity Reservation in any way, and they cannot view or modify instances that are owned by other consumers or the Capacity Reservation owner.

**Billing and Metering**

There are no additional charges for sharing Capacity Reservations.

The Capacity Reservation owner is billed for instances that they run inside the Capacity Reservation and for unused reserved capacity. Consumers are billed for the instances that they run inside the shared Capacity Reservation.

**Instance Limits**

All Capacity Reservation usage counts toward the Capacity Reservation owner's On-Demand Instance limits. This includes:

- Unused reserved capacity
- Usage by instances owned by the Capacity Reservation owner
- Usage by instances owned by consumers
Instances launched into the shared capacity by consumers count towards the Capacity Reservation owner's On-Demand Instance limit. Consumers' instance limits are a sum of their own On-Demand Instance limits and the capacity available in the shared Capacity Reservations to which they have access.

**Instance Lifecycle**

By working with Amazon EC2 to manage your instances from the moment you launch them through their termination, you ensure that your customers have the best possible experience with the applications or sites that you host on your instances.

The following illustration represents the transitions between instance states.

The following table provides a brief description of each instance state and indicates whether it is billed or not.

**Note**

The table indicates billing for instance usage only. Some AWS resources, such as Amazon EBS volumes and Elastic IP addresses, incur charges regardless of the instance's state. For more information, see *Avoiding Unexpected Charges* in the *AWS Billing and Cost Management User Guide*.

<table>
<thead>
<tr>
<th>Instance state</th>
<th>Description</th>
<th>Instance usage billing</th>
</tr>
</thead>
<tbody>
<tr>
<td>pending</td>
<td>The instance is preparing to enter the running state. An instance enters the pending state when it launches for the first time, or when it is restarted after being in the stopped state.</td>
<td>Not billed</td>
</tr>
<tr>
<td>running</td>
<td>The instance is running and ready for use.</td>
<td>Billed</td>
</tr>
<tr>
<td>Instance state</td>
<td>Description</td>
<td>Instance usage billing</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>stopping</td>
<td>The instance is preparing to be stopped or stop- hibernated.</td>
<td>Not billed if preparing to stop; Billed if preparing to hibernate</td>
</tr>
<tr>
<td>stopped</td>
<td>The instance is shut down and cannot be used. The instance can be restarted at any time.</td>
<td>Not billed</td>
</tr>
<tr>
<td>shutting down</td>
<td>The instance is preparing to be terminated.</td>
<td>Not billed</td>
</tr>
<tr>
<td>terminated</td>
<td>The instance has been permanently deleted and cannot be restarted.</td>
<td>Not billed</td>
</tr>
</tbody>
</table>

**Note**

Reserved Instances that applied to terminated instances are billed until the end of their term according to their payment option. For more information, see [Reserved Instances](#) (p. 197).

---

**Instance Launch**

When you launch an instance, it enters the pending state. The instance type that you specified at launch determines the hardware of the host computer for your instance. We use the Amazon Machine Image (AMI) you specified at launch to boot the instance. After the instance is ready for you, it enters the running state. You can connect to your running instance and use it the way that you’d use a computer sitting in front of you.

As soon as your instance transitions to the running state, you're billed for each hour or partial hour that you keep the instance running, even if the instance remains idle and you don't connect to it.

For more information, see [Launch Your Instance](#) (p. 333) and [Connecting to Your Windows Instance](#) (p. 379).

**Instance Stop and Start (Amazon EBS-Backed Instances Only)**

If your instance fails a status check or is not running your applications as expected, and if the root volume of your instance is an Amazon EBS volume, you can stop and start your instance to try to fix the problem.

When you stop your instance, it enters the stopping state, and then the stopped state. We don’t charge hourly usage or data transfer fees for your instance after you stop it, but we do charge for the storage for any Amazon EBS volumes. While your instance is in the stopped state, you can modify certain attributes of the instance, including the instance type.

When you start your instance, it enters the pending state, and in most cases, we move the instance to a new host computer. (Your instance may stay on the same host computer if there are no problems...
with the host computer. When you stop and start your instance, you lose any data on the instance store volumes on the previous host computer.

Your instance retains its private IPv4 address, which means that an Elastic IP address associated with the private IPv4 address or network interface is still associated with your instance. If your instance has an IPv6 address, it retains its IPv6 address.

Each time you transition an instance from stopped to running, we charge a full instance hour, even if these transitions happen multiple times within a single hour.

For more information, see Stop and Start Your Instance (p. 383).

Instance Hibernate (Amazon EBS-Backed Instances Only)

When you hibernate an instance, we signal the operating system to perform hibernation (suspend-to-disk), which saves the contents from the instance memory (RAM) to your Amazon EBS root volume. We persist the instance's Amazon EBS root volume and any attached Amazon EBS data volumes. When you restart your instance, the Amazon EBS root volume is restored to its previous state and the RAM contents are reloaded. Previously attached data volumes are reattached and the instance retains its instance ID.

When you hibernate your instance, it enters the stopping state, and then the stopped state. We don't charge hourly usage for a hibernated instance when it is in the stopped state, but we do charge while it is in the stopping state, unlike when you stop an instance (p. 330) without hibernating it. We don't charge usage for data transfer fees, but we do charge for the storage for any Amazon EBS volumes, including storage for the RAM data.

When you restart your hibernated instance, it enters the pending state, and in most cases, we move the instance to a new host computer. Your instance may stay on the same host computer if there are no problems with the host computer.

Your instance retains its private IPv4 address, which means that an Elastic IP address associated with the private IPv4 address or network interface is still associated with your instance. If your instance has an IPv6 address, it retains its IPv6 address.

For more information, see Hibernate Your Instance (p. 385).

Instance Reboot

You can reboot your instance using the Amazon EC2 console, a command line tool, and the Amazon EC2 API. We recommend that you use Amazon EC2 to reboot your instance instead of running the operating system reboot command from your instance.

Rebooting an instance is equivalent to rebooting an operating system. The instance remains on the same host computer and maintains its public DNS name, private IP address, and any data on its instance store volumes. It typically takes a few minutes for the reboot to complete, but the time it takes to reboot depends on the instance configuration.

Rebooting an instance doesn't start a new instance billing hour.

For more information, see Reboot Your Instance (p. 385).

Instance Retirement

An instance is scheduled to be retired when AWS detects the irreparable failure of the underlying hardware hosting the instance. When an instance reaches its scheduled retirement date, it is stopped or terminated by AWS. If your instance root device is an Amazon EBS volume, the instance is stopped, and
you can start it again at any time. If your instance root device is an instance store volume, the instance is
terminated, and cannot be used again.

For more information, see Instance Retirement (p. 386).

Instance Termination

When you've decided that you no longer need an instance, you can terminate it. As soon as the status of
an instance changes to shutting-down or terminated, you stop incurring charges for that instance.

If you enable termination protection, you can't terminate the instance using the console, CLI, or API.

After you terminate an instance, it remains visible in the console for a short while, and then the entry
is automatically deleted. You can also describe a terminated instance using the CLI and API. Resources
(such as tags) are gradually disassociated from the terminated instance, therefore may no longer be
visible on the terminated instance after a short while. You can't connect to or recover a terminated
instance.

Each Amazon EBS-backed instance supports the InstanceInitiatedShutdownBehavior attribute,
which controls whether the instance stops or terminates when you initiate shutdown from within the
instance itself. The default behavior is to stop the instance. You can modify the setting of this attribute
while the instance is running or stopped.

Each Amazon EBS volume supports the DeleteOnTermination attribute, which controls whether the
volume is deleted or preserved when you terminate the instance it is attached to. The default is to delete
the root device volume and preserve any other EBS volumes.

For more information, see Terminate Your Instance (p. 388).

Differences Between Reboot, Stop, Hibernate, and Terminate

The following table summarizes the key differences between rebooting, stopping, hibernating, and
terminating your instance.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Reboot</th>
<th>Stop/start (Amazon EBS-backed instances only)</th>
<th>Hibernate (Amazon EBS-backed instances only)</th>
<th>Terminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host computer</td>
<td>The instance stays on the same host computer</td>
<td>In most cases, we move the instance to a new host computer. Your instance may stay on the same host computer if there are no problems with the host computer.</td>
<td>In most cases, we move the instance to a new host computer. Your instance may stay on the same host computer if there are no problems with the host computer.</td>
<td>None</td>
</tr>
<tr>
<td>Private and public IPv4 addresses</td>
<td>These addresses stay the same</td>
<td>The instance keeps its private IPv4 address. The instance gets a new public IPv4 address, unless it has an Elastic IP address, which doesn't change during a stop/start.</td>
<td>The instance keeps its private IPv4 address. The instance gets a new public IPv4 address, unless it has an Elastic IP address, which doesn't change during a stop/start.</td>
<td>None</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Reboot</td>
<td>Stop/start (Amazon EBS-backed instances only)</td>
<td>Hibernate (Amazon EBS-backed instances only)</td>
<td>Terminate</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Elastic IP addresses (IPv4)</td>
<td>The Elastic IP address remains associated with the instance</td>
<td>The Elastic IP address remains associated with the instance</td>
<td>The Elastic IP address remains associated with the instance</td>
<td>The Elastic IP address is disassociated from the instance</td>
</tr>
<tr>
<td>IPv6 address</td>
<td>The address stays the same</td>
<td>The instance keeps its IPv6 address</td>
<td>The instance keeps its IPv6 address</td>
<td>None</td>
</tr>
<tr>
<td>Instance store volumes</td>
<td>The data is preserved</td>
<td>The data is erased</td>
<td>The data is erased</td>
<td>The data is erased</td>
</tr>
<tr>
<td>Root device volume</td>
<td>The volume is preserved</td>
<td>The volume is preserved</td>
<td>The volume is preserved</td>
<td>The volume is deleted by default</td>
</tr>
<tr>
<td>RAM (contents of memory)</td>
<td>The RAM is erased</td>
<td>The RAM is erased</td>
<td>The RAM is saved to a file on the root volume</td>
<td>The RAM is erased</td>
</tr>
<tr>
<td>Billing</td>
<td>The instance billing hour doesn't change.</td>
<td>You stop incurring charges for an instance as soon as its state changes to stopping. Each time an instance transitions from stopped to running, we start a new instance billing hour.</td>
<td>You incur charges while the instance is in the stopping state, but stop incurring charges when the instance is in the stopped state. Each time an instance transitions from stopped to running, we start a new instance billing hour.</td>
<td>You stop incurring charges for an instance as soon as its state changes to shutting-down.</td>
</tr>
</tbody>
</table>

Operating system shutdown commands always terminate an instance store-backed instance. You can control whether operating system shutdown commands stop or terminate an Amazon EBS-backed instance. For more information, see Changing the Instance Initiated Shutdown Behavior (p. 391).

## Launch Your Instance

An instance is a virtual server in the AWS Cloud. You launch an instance from an Amazon Machine Image (AMI). The AMI provides the operating system, application server, and applications for your instance.

When you sign up for AWS, you can get started with Amazon EC2 for free using the AWS Free Tier. You can use the free tier to launch and use a micro instance for free for 12 months. If you launch an instance that is not within the free tier, you incur the standard Amazon EC2 usage fees for the instance. For more information, see the Amazon EC2 Pricing.

You can launch an instance using the following methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Amazon EC2 console] Use the launch instance wizard to specify the launch parameters.</td>
<td>Launching an Instance Using the Launch Instance Wizard (p. 334)</td>
</tr>
<tr>
<td>Method</td>
<td>Documentation</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>[Amazon EC2 console] Create a launch template and launch the instance from the launch template.</td>
<td>Launching an Instance from a Launch Template (p. 338)</td>
</tr>
<tr>
<td>[Amazon EC2 console] Use an existing instance as the base.</td>
<td>Launching an Instance Using Parameters from an Existing Instance (p. 348)</td>
</tr>
<tr>
<td>[Amazon EC2 console] Use an AMI that you purchased from the AWS Marketplace.</td>
<td>Launching an AWS Marketplace Instance (p. 350)</td>
</tr>
<tr>
<td>[AWS CLI] Use an AMI that you select.</td>
<td>Using Amazon EC2 through the AWS CLI</td>
</tr>
<tr>
<td>[AWS CLI] Use EC2 Fleet to provision capacity across different EC2 instance types and Availability Zones, and across On-Demand Instance, Reserved Instance, and Spot Instance purchase models.</td>
<td>Launching an EC2 Fleet (p. 351)</td>
</tr>
</tbody>
</table>

After you launch your instance, you can connect to it and use it. To begin, the instance state is pending. When the instance state is running, the instance has started booting. There might be a short time before you can connect to the instance. The instance receives a public DNS name that you can use to contact the instance from the internet. The instance also receives a private DNS name that other instances within the same VPC can use to contact the instance. For more information about connecting to your instance, see Connecting to Your Windows Instance (p. 379).

When you are finished with an instance, be sure to terminate it. For more information, see Terminate Your Instance (p. 388).

**Launching an Instance Using the Launch Instance Wizard**

Before you launch your instance, be sure that you are set up. For more information, see Setting Up with Amazon EC2 (p. 12).

**Important**

When you launch an instance that's not within the AWS Free Tier, you are charged for the time that the instance is running, even if it remains idle.

**Launching Your Instance from an AMI**

When you launch an instance, you must select a configuration, known as an Amazon Machine Image (AMI). An AMI contains the information required to create a new instance. For example, an AMI might contain the software required to act as a web server: for example, Windows, Apache, and your website.

**Tip**

To ensure faster instance launches, break up large requests into smaller batches. For example, create five separate launch requests for 100 instances each instead of one launch request for 500 instances.

**To launch an instance**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation bar at the top of the screen, the current Region is displayed (for example, US East (Ohio)). Select a Region for the instance that meets your needs. This choice is important
because some Amazon EC2 resources can be shared between Regions, while others can’t. For more information, see Resource Locations (p. 955).

3. From the Amazon EC2 console dashboard, choose **Launch Instance**.

4. On the **Choose an Amazon Machine Image (AMI)** page, choose an AMI as follows:

   a. Select the type of AMI to use in the left pane:

      **Quick Start**

      A selection of popular AMIs to help you get started quickly. To select an AMI that is eligible for the free tier, choose **Free tier only** in the left pane. These AMIs are marked **Free tier eligible**.

      **My AMIs**

      The private AMIs that you own, or private AMIs that have been shared with you. To view AMIs shared with you, choose **Shared with me** in the left pane.

      **AWS Marketplace**

      An online store where you can buy software that runs on AWS, including AMIs. For more information about launching an instance from the AWS Marketplace, see Launching an AWS Marketplace Instance (p. 350).

      **Community AMIs**

      The AMIs that AWS community members have made available for others to use. To filter the list of AMIs by operating system, choose the appropriate check box under **Operating system**. You can also filter by architecture and root device type.

   b. Check the **Virtualization type** listed for each AMI. Notice which AMIs are the type that you need, either **hvm** or **paravirtual**. For example, some instance types require HVM.

   c. Choose an AMI that meets your needs, and then choose **Select**.

5. On the **Choose an Instance Type** page, select the hardware configuration and size of the instance to launch. Larger instance types have more CPU and memory. For more information, see Instance Types (p. 119).

   To remain eligible for the free tier, choose the **t2.micro** instance type. For more information, see Burstable Performance Instances (p. 132).

   By default, the wizard displays current generation instance types, and selects the first available instance type based on the AMI that you selected. To view previous generation instance types, choose **All generations** from the filter list.

   **Note**

   To set up an instance quickly for testing purposes, choose **Review and Launch** to accept the default configuration settings, and launch your instance. Otherwise, to configure your instance further, choose **Next: Configure Instance Details**.

6. On the **Configure Instance Details** page, change the following settings as necessary (expand **Advanced Details** to see all the settings), and then choose **Next: Add Storage**:

   - **Number of instances**: Enter the number of instances to launch.

   - **Purchasing option**: Choose **Request Spot instances** to launch a Spot Instance. This adds and removes options from this page. Set your maximum price, and optionally update the request type,
interruption behavior, and request validity. For more information, see Creating a Spot Instance Request (p. 248).

- **Network**: Select the VPC, or to create a new VPC, choose Create new VPC to go to the Amazon VPC console. When you have finished, return to the wizard and choose Refresh to load your VPC in the list.

- **Subnet**: Select the subnet into which to launch your instance. You can select No preference to let AWS choose a default subnet in any Availability Zone. To create a new subnet, choose Create new subnet to go to the Amazon VPC console. When you are done, return to the wizard and choose Refresh to load your subnet in the list.

- **Auto-assign Public IP**: Specify whether your instance receives a public IPv4 address. By default, instances in a default subnet receive a public IPv4 address and instances in a nondefault subnet do not. You can select Enable or Disable to override the subnet's default setting. For more information, see Public IPv4 Addresses and External DNS Hostnames (p. 657).

- **Auto-assign IPv6 IP**: Specify whether your instance receives an IPv6 address from the range of the subnet. Select Enable or Disable to override the subnet's default setting. This option is only available if you've associated an IPv6 CIDR block with your VPC and subnet. For more information, see Your VPC and Subnets in the Amazon VPC User Guide.

- **Domain join directory**: Select the AWS Directory Service directory (domain) to which your Windows instance is joined. The directory must be in the same VPC that you selected for your instance. If you select a domain, you must select an IAM role. For more information, see Launching an Instance (Simple AD and Microsoft AD).

- **Capacity Reservation**: Specify whether to launch the instance into shared capacity or an existing Capacity Reservation. For more information, see Launching Instances into an Existing Capacity Reservation (p. 321).

- **IAM role**: Select an AWS Identity and Access Management (IAM) role to associate with the instance. For more information, see IAM Roles for Amazon EC2 (p. 646).

- **CPU options**: Choose Specify CPU options to specify a custom number of vCPUs during launch. Set the number of CPU cores and threads per core. For more information, see Optimizing CPU Options (p. 447).

- **Shutdown behavior**: Select whether the instance should stop or terminate when shut down. For more information, see Changing the Instance Initiated Shutdown Behavior (p. 391).

- **Enable termination protection**: To prevent accidental termination, select this check box. For more information, see Enabling Termination Protection for an Instance (p. 390).

- **Monitoring**: Select this check box to enable detailed monitoring of your instance using Amazon CloudWatch. Additional charges apply. For more information, see Monitoring Your Instances Using CloudWatch (p. 546).

- **EBS-Optimized instance**: An Amazon EBS-optimized instance uses an optimized configuration stack and provides additional, dedicated capacity for Amazon EBS I/O. If the instance type supports this feature, select this check box to enable it. Additional charges apply. For more information, see Amazon EBS–Optimized Instances (p. 876).

- **Tenancy**: If you are launching your instance into a VPC, you can choose to run your instance on isolated, dedicated hardware (Dedicated) or on a Dedicated Host (Dedicated host). Additional charges may apply. For more information, see Dedicated Instances (p. 312) and Dedicated Hosts (p. 294).

- **T2/T3 Unlimited**: Select this check box to enable applications to burst beyond the baseline for as long as needed. Additional charges may apply. For more information, see Burstable Performance Instances (p. 132).

- **Network interfaces**: If you selected a specific subnet, you can specify up to two network interfaces for your instance:

  - For **Network Interface**, select New network interface to let AWS create a new interface, or select an existing, available network interface.

  - For **Primary IP**, enter a private IPv4 address from the range of your subnet, or leave Auto-assign to let AWS choose a private IPv4 address for you.
• For **Secondary IP addresses**, choose **Add IP** to assign more than one private IPv4 address to the selected network interface.

• (IPv6-only) For **IPv6 IPs**, choose **Add IP**, and enter an IPv6 address from the range of the subnet, or leave **Auto-assign** to let AWS choose one for you.

• Choose **Add Device** to add a secondary network interface. A secondary network interface can reside in a different subnet of the VPC, provided it's in the same Availability Zone as your instance.

For more information, see Elastic Network Interfaces (p. 679). If you specify more than one network interface, your instance cannot receive a public IPv4 address. Additionally, if you specify an existing network interface for eth0, you cannot override the subnet's public IPv4 setting using **Auto-assign Public IP**. For more information, see Assigning a Public IPv4 Address During Instance Launch (p. 661).

• **Kernel ID**: (Only valid for paravirtual (PV) AMIs) Select **Use default** unless you want to use a specific kernel.

• **RAM disk ID**: (Only valid for paravirtual (PV) AMIs) Select **Use default** unless you want to use a specific RAM disk. If you have selected a kernel, you may need to select a specific RAM disk with the drivers to support it.

• **Placement group**: A placement group determines the placement strategy of your instances. Select an existing placement group, or create a new one. This option is only available if you've selected an instance type that supports placement groups. For more information, see Placement Groups (p. 710).

• **User data**: You can specify user data to configure an instance during launch, or to run a configuration script. To attach a file, select the **As file** option and browse for the file to attach.

7. The AMI you selected includes one or more volumes of storage, including the root device volume. On the **Add Storage** page, you can specify additional volumes to attach to the instance by choosing **Add New Volume**. Configure each volume as follows, and then choose **Next: Add Tags**.

• **Type**: Select instance store or Amazon EBS volumes to associate with your instance. The types of volume available in the list depend on the instance type you've chosen. For more information, see Amazon EC2 Instance Store (p. 917) and Amazon EBS Volumes (p. 783).

• **Device**: Select from the list of available device names for the volume.

• **Snapshot**: Enter the name or ID of the snapshot from which to restore a volume. You can also search for available shared and public snapshots by typing text into the **Snapshot** field. Snapshot descriptions are case-sensitive.

• **Size**: For EBS volumes, you can specify a storage size. Even if you have selected an AMI and instance that are eligible for the free tier, to stay within the free tier, you must stay under 30 GiB of total storage. For more information, see Constraints on the Size and Configuration of an EBS Volume (p. 798).

• **Volume Type**: For EBS volumes, select a volume type. For more information, see Amazon EBS Volume Types (p. 786).

• **IOPS**: If you have selected a Provisioned IOPS SSD volume type, then you can enter the number of I/O operations per second (IOPS) that the volume can support.

• **Delete on Termination**: For Amazon EBS volumes, select this check box to delete the volume when the instance is terminated. For more information, see Preserving Amazon EBS Volumes on Instance Termination (p. 392).

• **Encrypted**: If the instance type supports EBS encryption, you can specify the encryption state of the volume. If you have enabled encryption by default in this Region, the default CMK is selected for you. You can select a different key or disable encryption. For more information, see Amazon EBS Encryption (p. 865).

8. On the **Add Tags** page, specify **tags** (p. 964) by providing key and value combinations. You can tag the instance, the volumes, or both. For Spot Instances, you can tag the Spot Instance request
only. Choose **Add another tag** to add more than one tag to your resources. Choose **Next: Configure Security Group** when you are done.

9. On the **Configure Security Group** page, use a security group to define firewall rules for your instance. These rules specify which incoming network traffic is delivered to your instance. All other traffic is ignored. (For more information about security groups, see *Amazon EC2 Security Groups for Windows Instances (p. 583)*.) Select or create a security group as follows, and then choose **Review and Launch.**

   a. To select an existing security group, choose **Select an existing security group**, and select your security group. You can't edit the rules of an existing security group, but you can copy them to a new group by choosing **Copy to new**. Then you can add rules as described in the next step.

   b. To create a new security group, choose **Create a new security group**. The wizard automatically defines the launch-wizard-x security group and creates an inbound rule to allow you to connect to your instance over RDP (port 3389).

   c. You can add rules to suit your needs. For example, if your instance is a web server, open ports 80 (HTTP) and 443 (HTTPS) to allow internet traffic.

   To add a rule, choose **Add Rule**, select the protocol to open to network traffic, and then specify the source. Choose **My IP** from the **Source** list to let the wizard add your computer's public IP address. However, if you are connecting through an ISP or from behind your firewall without a static IP address, you need to find out the range of IP addresses used by client computers.

   **Warning**
   Rules that enable all IP addresses (0.0.0.0/0) to access your instance over SSH or RDP are acceptable for this short exercise, but are unsafe for production environments. You should authorize only a specific IP address or range of addresses to access your instance.

10. On the **Review Instance Launch** page, check the details of your instance, and make any necessary changes by choosing the appropriate **Edit** link.

    When you are ready, choose **Launch**.

11. In the **Select an existing key pair or create a new key pair** dialog box, you can choose an existing key pair, or create a new one. For example, choose **Choose an existing key pair**, then select the key pair you created when getting set up.

    To launch your instance, select the acknowledgment check box, then choose **Launch Instances**.

    **Important**
    If you choose the **Proceed without key pair** option, you won't be able to connect to the instance unless you choose an AMI that is configured to allow users another way to log in.

12. (Optional) You can create a status check alarm for the instance (additional fees may apply). (If you're not sure, you can always add one later.) On the confirmation screen, choose **Create status check alarms** and follow the directions. For more information, see *Creating and Editing Status Check Alarms (p. 538)*.

13. If the instance fails to launch or the state immediately goes to **terminated** instead of **running**, see *Troubleshooting Instance Launch Issues (p. 1042)*.

**Launching an Instance from a Launch Template**

You can create a *launch template* that contains the configuration information to launch an instance. Launch templates enable you to store launch parameters so that you do not have to specify them every time you launch an instance. For example, a launch template can contain the AMI ID, instance type, and network settings that you typically use to launch instances. When you launch an instance using the Amazon EC2 console, an AWS SDK, or a command line tool, you can specify the launch template to use.
For each launch template, you can create one or more numbered *launch template versions*. Each version can have different launch parameters. When you launch an instance from a launch template, you can use any version of the launch template. If you do not specify a version, the default version is used. You can set any version of the launch template as the default version—by default, it's the first version of the launch template.

The following diagram shows a launch template with three versions. The first version specifies the instance type, AMI ID, subnet, and key pair to use to launch the instance. The second version is based on the first version and also specifies a security group for the instance. The third version uses different values for some of the parameters. Version 2 is set as the default version. If you launched an instance from this launch template, the launch parameters from version 2 would be used if no other version were specified.

### Contents
- Launch Template Restrictions (p. 339)
- Using Launch Templates to Control Launch Parameters (p. 340)
- Controlling the Use of Launch Templates (p. 340)
- Creating a Launch Template (p. 340)
- Managing Launch Template Versions (p. 345)
- Launching an Instance from a Launch Template (p. 346)
- Using Launch Templates with Amazon EC2 Auto Scaling (p. 347)
- Using Launch Templates with EC2 Fleet (p. 348)
- Using Launch Templates with Spot Fleet (p. 348)
- Deleting a Launch Template (p. 348)

### Launch Template Restrictions

The following rules apply to launch templates and launch template versions:

- You are limited to creating 5,000 launch templates per Region and 10,000 versions per launch template.
- Launch parameters are optional. However, you must ensure that your request to launch an instance includes all required parameters. For example, if your launch template does not include an AMI ID, you must specify both the launch template and an AMI ID when you launch an instance.
- Launch template parameters are not validated when you create the launch template. Ensure that you specify the correct values for the parameters and that you use supported parameter combinations. For example, to launch an instance in a placement group, you must specify a supported instance type.
- You can tag a launch template, but you cannot tag a launch template version.
- Launch template versions are numbered in the order in which they are created. When you create a launch template version, you cannot specify the version number yourself.
Using Launch Templates to Control Launch Parameters

A launch template can contain all or some of the parameters to launch an instance. When you launch an instance using a launch template, you can override parameters that are specified in the launch template. Or, you can specify additional parameters that are not in the launch template.

**Note**
You cannot remove launch template parameters during launch (for example, you cannot specify a null value for the parameter). To remove a parameter, create a new version of the launch template without the parameter and use that version to launch the instance.

To launch instances, IAM users must have permissions to use the `ec2:RunInstances` action. You must also have permissions to create or use the resources that are created or associated with the instance. You can use resource-level permissions for the `ec2:RunInstances` action to control the launch parameters that users can specify. Alternatively, you can grant users permissions to launch an instance using a launch template. This enables you to manage launch parameters in a launch template rather than in an IAM policy, and to use a launch template as an authorization vehicle for launching instances. For example, you can specify that users can only launch instances using a launch template, and that they can only use a specific launch template. You can also control the launch parameters that users can override in the launch template. For example policies, see Launch Templates (p. 631).

Controlling the Use of Launch Templates

By default, IAM users do not have permissions to work with launch templates. You can create an IAM user policy that grants users permissions to create, modify, describe, and delete launch templates and launch template versions. You can also apply resource-level permissions to some launch template actions to control a user's ability to use specific resources for those actions. For more information, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608) and the following example policies: Example: Working with Launch Templates (p. 638).

Take care when granting users permissions to use the `ec2:CreateLaunchTemplate` and `ec2:CreateLaunchTemplateVersion` actions. These actions do not support resource-level permissions that enable you to control which resources users can specify in the launch template. To restrict the resources that are used to launch an instance, ensure that you grant permissions to create launch templates and launch template versions only to appropriate administrators.

Creating a Launch Template

Create a new launch template using parameters that you define, or use an existing launch template or an instance as the basis for a new launch template.

**To create a new launch template using defined parameters (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Launch Templates, Create launch template.
3. For **Launch template name**, enter a descriptive name for the launch template. To tag the launch template on creation, choose Show Tags, Add Tag, and then enter a tag key and value pair.
4. For **Template version description**, provide a brief description of the launch template version.
5. For **Launch template contents**, provide the following information:
   - **AMI ID**: An AMI from which to launch the instance. To search through all available AMIs, choose Search for AMI. To select a commonly used AMI, choose Quick Start. Or, choose AWS Marketplace or Community AMIs. You can use an AMI that you own or find a suitable AMI (p. 51).
   - **Instance type**: Ensure that the instance type is compatible with the AMI that you've specified. For more information, see Instance Types (p. 119).
   - **Key pair name**: The key pair for the instance. For more information, see Amazon EC2 Key Pairs and Windows Instances (p. 578).
• **Network type**: If applicable, whether to launch the instance into a VPC or EC2-Classic. If you choose VPC, specify the subnet in the **Network interfaces** section. If you choose Classic, ensure that the specified instance type is supported in EC2-Classic and specify the Availability Zone for the instance.

• **Security Groups**: One or more security groups to associate with the instance. For more information, see Amazon EC2 Security Groups for Windows Instances (p. 583).

6. For **Network interfaces**, you can specify up to two network interfaces (p. 679) for the instance.

• **Device**: The device number for the network interface, for example, `eth0` for the primary network interface. If you leave the field blank, AWS creates the primary network interface.

• **Network interface**: The ID of the network interface, or leave blank to let AWS create a new network interface.

• **Description**: (Optional) A description for the new network interface.

• **Subnet**: The subnet in which to create a new network interface. For the primary network interface (`eth0`), this is the subnet in which the instance is launched. If you've entered an existing network interface for `eth0`, the instance is launched in the subnet in which the network interface is located.

• **Auto-assign public IP**: Whether to automatically assign a public IP address to the network interface with the device index of `eth0`. This setting can only be enabled for a single, new network interface.

• **Primary IP**: A private IPv4 address from the range of your subnet. Leave blank to let AWS choose a private IPv4 address for you.

• **Secondary IP**: A secondary private IPv4 address from the range of your subnet. Leave blank to let AWS choose one for you.

• **(IPv6-only) IPv6 IPs**: An IPv6 address from the range of the subnet.

• **Security group ID**: The ID of a security group in your VPC with which to associate the network interface.

• **Delete on termination**: Whether the network interface is deleted when the instance is deleted.

7. For **Storage (Volumes)**, specify volumes to attach to the instance besides the volumes specified by the AMI.

• **Volume type**: The instance store or Amazon EBS volumes with which to associate your instance. The type of volume depends on the instance type that you've chosen. For more information, see Amazon EC2 Instance Store (p. 917) and Amazon EBS Volumes (p. 783).

• **Device name**: A device name for the volume.

• **Snapshot**: The ID of the snapshot from which to create the volume.

• **Size**: For Amazon EBS volumes, the storage size.

• **Volume type**: For Amazon EBS volumes, the volume type. For more information, see Amazon EBS Volume Types (p. 786).

• **IOPS**: For the Provisioned IOPS SSD volume type, the number of I/O operations per second (IOPS) that the volume can support.

• **Delete on termination**: For Amazon EBS volumes, whether to delete the volume when the instance is terminated. For more information, see Preserving Amazon EBS Volumes on Instance Termination (p. 392).

• **Encrypted**: If the instance type supports EBS encryption, you can enable encryption for the volume. If you have enabled encryption by default in this Region, encryption is enabled for you. For more information, see Amazon EBS Encryption (p. 865).

• **Key**: The CMK to use for EBS encryption. You can specify the ARN of any customer master key (CMK) that you created using the AWS Key Management Service. If you specify a CMK, you must also use Encrypted to enable encryption.

8. For **Tags**, specify tags (p. 964) by providing key and value combinations. You can tag the instance, the volumes, or both.
9. For **Advanced Details**, expand the section to view the fields and specify any additional parameters for the instance.

   - **Purchasing option**: The purchasing model. Choose **Request Spot instances** to request Spot Instances at the Spot price, capped at the On-Demand price, and choose **Customize Spot parameters** to change the default Spot Instance settings. If you do not request a Spot Instance, EC2 launches an On-Demand Instance by default. For more information, see **Spot Instances (p. 231)**.
   - **IAM instance profile**: An AWS Identity and Access Management (IAM) instance profile to associate with the instance. For more information, see **IAM Roles for Amazon EC2 (p. 646)**.
   - **Shutdown behavior**: Whether the instance should stop or terminate when shut down. For more information, see **Changing the Instance Initiated Shutdown Behavior (p. 391)**.
   - **Stop - Hibernate behavior**: Whether the instance is enabled for hibernation. This field is only valid for instances that meet the hibernation prerequisites. For more information, see **Hibernate Your Instance (p. 385)**.
   - **Termination protection**: Whether to prevent accidental termination. For more information, see **Enabling Termination Protection for an Instance (p. 390)**.
   - **Monitoring**: Whether to enable detailed monitoring of the instance using Amazon CloudWatch. Additional charges apply. For more information, see **Monitoring Your Instances Using CloudWatch (p. 546)**.
   - **Elastic Graphics**: An Elastic Graphics accelerator to attach to the instance. Not all instance types support Elastic Graphics. For more information, see **Amazon Elastic Graphics (p. 520)**.
   - **T2/T3 Unlimited**: Whether to enable applications to burst beyond the baseline for as long as needed. This field is only valid for T2 and T3 instances. Additional charges may apply. For more information, see **Burstable Performance Instances (p. 132)**.
   - **Placement group name**: Specify a placement group in which to launch the instance. Not all instance types can be launched in a placement group. For more information, see **Placement Groups (p. 710)**.
   - **EBS-optimized instance**: Provides additional, dedicated capacity for Amazon EBS I/O. Not all instance types support this feature, and additional charges apply. For more information, see **Amazon EBS–Optimized Instances (p. 876)**.
   - **Tenancy**: Choose whether to run your instance on shared hardware (Shared), isolated, dedicated hardware (Dedicated), or on a Dedicated Host (Dedicated host). Additional charges may apply. For more information, see **Dedicated Instances (p. 312)** and **Dedicated Hosts (p. 294)**. If you specify a Dedicated Host, you can choose a specific host and the affinity for the instance.
   - **RAM disk ID**: A RAM disk for the instance. If you have specified a kernel, you may need to specify a specific RAM disk with the drivers to support it. Only valid for paravirtual (PV) AMIs.
   - **Kernel ID**: A kernel for the instance. Only valid for paravirtual (PV) AMIs.
   - **User data**: You can specify user data to configure an instance during launch, or to run a configuration script. For more information, see **Running Commands on Your Windows Instance at Launch (p. 472)**.

10. Choose **Create launch template**.

### To create a launch template from an existing launch template (console)

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Launch Templates**.
3. Choose **Create launch template**. Provide a name, description, and tags for the launch template.
4. For **Source template**, choose a launch template on which to base the new launch template.
5. For **Source template version**, choose the launch template version on which to base the new launch template.
6. Adjust any launch parameters as required, and choose **Create launch template**.
To create a launch template from an instance (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance, and choose Actions, Create Template From Instance.
4. Provide a name, description, and tags, and adjust the launch parameters as required.

   **Note**
   When you create a launch template from an instance, the instance's network interface IDs and IP addresses are not included in the template.
5. Choose Create Template From Instance.

To create a launch template (AWS CLI)

- Use the create-launch-template (AWS CLI) command. The following example creates a launch template that specifies the following:
  - A tag for the launch template (purpose=production)
  - The instance type (r4.4xlarge) and AMI (ami-8c1be5f6) to launch
  - The number of cores (4) and threads per core (2) for a total of 8 vCPUs (4 cores x 2 threads)
  - The subnet in which to launch the instance (subnet-7b16de0c)

The template assigns a public IP address and an IPv6 address to the instance and creates a tag for the instance(Name=webserver).

```
aws ec2 create-launch-template --launch-template-name TemplateForWebServer
--version-description WebVersion1 --tag-specifications 'ResourceType=launch-template,Tags=[{Key=purpose,Value=production}]' --launch-template-data file://template-data.json
```

The following is an example template-data.json file:

```json
{
    "NetworkInterfaces": [{
        "AssociatePublicIpAddress": true,
        "DeviceIndex": 0,
        "Ipv6AddressCount": 1,
        "SubnetId": "subnet-7b16de0c"
    }],
    "ImageId": "ami-8c1be5f6",
    "InstanceType": "r4.4xlarge",
    "TagSpecifications": [{
        "ResourceType": "instance",
        "Tags": [{
            "Key": "Name",
            "Value": "webserver"
        }]
    }],
    "CpuOptions": {
        "CoreCount": 4,
        "ThreadsPerCore": 2
    }
}
```

The following is example output.
To get instance data for a launch template (AWS CLI)

- Use the `get-launch-template-data` (AWS CLI) command and specify the instance ID. You can use the output as a base to create a new launch template or launch template version. By default, the output includes a top-level `LaunchTemplateData` object, which cannot be specified in your launch template data. Use the `--query` option to exclude this object.

```
aws ec2 get-launch-template-data --instance-id i-0123d646e8048babc --query "LaunchTemplateData"
```

The following is example output.

```
{
   "Monitoring": {},
   "ImageId": "ami-8c1be5f6",
   "BlockDeviceMappings": [
      {
         "DeviceName": "/dev/xvda",
         "Ebs": {
            "DeleteOnTermination": true
         }
      }
   ],
   "EbsOptimized": false,
   "Placement": {
      "Tenancy": "default",
      "GroupName": ":",
      "AvailabilityZone": "us-east-1a"
   },
   "InstanceType": "t2.micro",
   "NetworkInterfaces": [
      {
         "Description": "",
         "NetworkInterfaceId": "eni-35306abc",
         "PrivateIpAddresses": [
            {
               "Primary": true,
               "PrivateIpAddress": "10.0.0.72"
            }
         ],
         "SubnetId": "subnet-7b16de0c",
         "Groups": [
            "sg-7c27019"
         ],
         "Ipv6Addresses": [
            {
               "Ipv6Address": "2001:db8:1234:1a00::123"
            }
         ]
      }
   ]
}
```
"PrivateIpAddress": "10.0.0.72"
}]
}

You can write the output directly to a file, for example:

```
aws ec2 get-launch-template-data --instance-id i-0123d646e8048babc --query "LaunchTemplateData" >> instance-data.json
```

Managing Launch Template Versions

You can create launch template versions for a specific launch template, set the default version, and delete versions that you no longer require.

**Tasks**

- Creating a Launch Template Version (p. 345)
- Setting the Default Launch Template Version (p. 346)
- Deleting a Launch Template Version (p. 346)

Creating a Launch Template Version

When you create a launch template version, you can specify new launch parameters or use an existing version as the base for the new version. For more information about the launch parameters, see Creating a Launch Template (p. 340).

**To create a launch template version (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Launch Templates.
3. Choose Create launch template.
4. For What would you like to do, choose Create a new template version
5. For Launch template name, select the name of the existing launch template from the list.
6. For Template version description, type a description for the launch template version.
7. (Optional) Select a version of the launch template, or a version of a different launch template, to use as a base for the new launch template version. The new launch template version inherits the launch parameters from this launch template version.
8. Modify the launch parameters as required, and choose Create launch template.

**To create a launch template version (AWS CLI)**

- Use the create-launch-template-version (AWS CLI) command. You can specify a source version on which to base the new version. The new version inherits the launch parameters from this version, and you can override parameters using --launch-template-data. The following example creates a new version based on version 1 of the launch template and specifies a different AMI ID.

```
aws ec2 create-launch-template-version --launch-template-id lt-0abcd290751193123 --version-description WebVersion2 --source-version 1 --launch-template-data "ImageId=ami-c998b6b2"
```
Amazon Elastic Compute Cloud
User Guide for Windows Instances
Launch

Setting the Default Launch Template Version

You can set the default version for the launch template. When you launch an instance from a launch template and do not specify a version, the instance is launched using the parameters of the default version.

To set the default launch template version (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Launch Templates.
3. Select the launch template and choose Actions, Set default version.
4. For Default version, select the version number and choose Set as default version.

To set the default launch template version (AWS CLI)

- Use the modify-launch-template (AWS CLI) command and specify the version that you want to set as the default.

```bash
aws ec2 modify-launch-template --launch-template-id lt-0abcd290751193123 --default-version 2
```

Deleting a Launch Template Version

If you no longer require a launch template version, you can delete it. You cannot replace the version number after you delete it. You cannot delete the default version of the launch template; you must first assign a different version as the default.

To delete a launch template version (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Launch Templates.
3. Select the launch template and choose Actions, Delete template version.
4. Select the version to delete and choose Delete launch template version.

To delete a launch template version (AWS CLI)

- Use the delete-launch-template-versions (AWS CLI) command and specify the version numbers to delete.

```bash
aws ec2 delete-launch-template-versions --launch-template-id lt-0abcd290751193123 --versions 1
```

Launching an Instance from a Launch Template

You can use the parameters contained in a launch template to launch an instance. You have the option to override or add launch parameters before you launch the instance.

Instances that are launched using a launch template are automatically assigned two tags with the keys aws:ec2launchtemplate:id and aws:ec2launchtemplate:version. You cannot remove or edit these tags.

To launch an instance from a launch template (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Launch Templates**.
3. Select the launch template and choose **Actions, Launch instance from template**.
4. Select the launch template version to use.
5. (Optional) You can override or add launch template parameters by changing and adding parameters in the **Instance details** section.
6. Choose **Launch instance from template**.

**To launch an instance from a launch template (AWS CLI)**

- Use the `run-instances` AWS CLI command and specify the `--launch-template` parameter. Optionally specify the launch template version to use. If you don't specify the version, the default version is used.

  ```sh
daemon run-instances --launch-template LaunchTemplateId=lt-0abcd290751193123,Version=1
  ```

- To override a launch template parameter, specify the parameter in the `run-instances` command. The following example overrides the instance type that's specified in the launch template (if any).

  ```sh
daemon run-instances --launch-template LaunchTemplateId=lt-0abcd290751193123 --instance-type t2.small
  ```

- If you specify a nested parameter that's part of a complex structure, the instance is launched using the complex structure as specified in the launch template plus any additional nested parameters that you specify.

  In the following example, the instance is launched with the tag `Owner=TeamA` as well as any other tags that are specified in the launch template. If the launch template has an existing tag with a key of `Owner`, the value is replaced with `TeamA`.

  ```sh
daemon run-instances --launch-template LaunchTemplateId=lt-0abcd290751193123 --tag-specifications "ResourceType=instance,Tags=[{Key=Owner,Value=TeamA}]"
  ```

  In the following example, the instance is launched with a volume with the device name `/dev/xvdb` as well as any other block device mappings that are specified in the launch template. If the launch template has an existing volume defined for `/dev/xvdb`, its values are replaced with specified values.

  ```sh
daemon run-instances --launch-template LaunchTemplateId=lt-0abcd290751193123 --block-device-mappings "DeviceName=/dev/xvdb,Ebs={VolumeSize=20,VolumeType=gp2}"
  ```

If the instance fails to launch or the state immediately goes to terminated instead of running, see **Troubleshooting Instance Launch Issues (p. 1042)**.

**Using Launch Templates with Amazon EC2 Auto Scaling**

You can create an Auto Scaling group and specify a launch template to use for the group. When Amazon EC2 Auto Scaling launches instances in the Auto Scaling group, it uses the launch parameters defined in the associated launch template.

For more information, see **Creating an Auto Scaling Group Using a Launch Template** in the *Amazon EC2 Auto Scaling User Guide*.

**To create or update an Amazon EC2 Auto Scaling group with a launch template (AWS CLI)**

- Use the `create-auto-scaling-group` or the `update-auto-scaling-group` AWS CLI command and specify the `--launch-template` parameter.
Using Launch Templates with EC2 Fleet

You can create an EC2 Fleet request and specify a launch template in the instance configuration. When Amazon EC2 fulfills the EC2 Fleet request, it uses the launch parameters defined in the associated launch template. You can override some of the parameters that are specified in the launch template.

For more information, see Creating an EC2 Fleet (p. 370).

To create an EC2 Fleet with a launch template (AWS CLI)

- Use the create-fleet AWS CLI command. Use the --launch-template-configs parameter to specify the launch template and any overrides for the launch template.

Using Launch Templates with Spot Fleet

You can create a Spot Fleet request and specify a launch template in the instance configuration. When Amazon EC2 fulfills the Spot Fleet request, it uses the launch parameters defined in the associated launch template. You can override some of the parameters that are specified in the launch template.

For more information, see Spot Fleet Requests (p. 254).

To create a Spot Fleet request with a launch template (AWS CLI)

- Use the request-spot-fleet AWS CLI command. Use the LaunchTemplateConfigs parameter to specify the launch template and any overrides for the launch template.

Deleting a Launch Template

If you no longer require a launch template, you can delete it. Deleting a launch template deletes all of its versions.

To delete a launch template (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Launch Templates.
3. Select the launch template and choose Actions, Delete template.
4. Choose Delete launch template.

To delete a launch template (AWS CLI)

- Use the delete-launch-template (AWS CLI) command and specify the launch template.

```bash
aws ec2 delete-launch-template --launch-template-id lt-01238c059e3466abc
```

Launching an Instance Using Parameters from an Existing Instance

The Amazon EC2 console provides a Launch More Like This wizard option that enables you to use a current instance as a base for launching other instances. This option automatically populates the Amazon EC2 launch wizard with certain configuration details from the selected instance.
Note
The Launch More Like This wizard option does not clone your selected instance; it only replicates some configuration details. To create a copy of your instance, first create an AMI from it, then launch more instances from the AMI. Alternatively, create a launch template (p. 338) to store the launch parameters for your instances.

The following configuration details are copied from the selected instance into the launch wizard:

- AMI ID
- Instance type
- Availability Zone, or the VPC and subnet in which the selected instance is located
- Public IPv4 address. If the selected instance currently has a public IPv4 address, the new instance receives a public IPv4 address - regardless of the selected instance's default public IPv4 address setting. For more information about public IPv4 addresses, see Public IPv4 Addresses and External DNS Hostnames (p. 657).
- Placement group, if applicable
- IAM role associated with the instance, if applicable
- Shutdown behavior setting (stop or terminate)
- Termination protection setting (true or false)
- CloudWatch monitoring (enabled or disabled)
- Amazon EBS-optimization setting (true or false)
- Tenancy setting, if launching into a VPC (shared or dedicated)
- Kernel ID and RAM disk ID, if applicable
- User data, if specified
- Tags associated with the instance, if applicable
- Security groups associated with the instance
- Association information. If the selected instance is associated with a configuration file, the same file is automatically associated with the new instance. If the configuration file includes a joined domain configuration, the new instance is joined to the same domain. For more information about joining a domain, see Launching an Instance (Simple AD and Microsoft AD).

The following configuration details are not copied from your selected instance; instead, the wizard applies their default settings or behavior:

- Number of network interfaces: The default is one network interface, which is the primary network interface (eth0).
- Storage: The default storage configuration is determined by the AMI and the instance type.

To use your current instance as a template

1. On the Instances page, select the instance you want to use.
2. Choose Actions, and then Launch More Like This.
3. The launch wizard opens on the Review Instance Launch page. You can check the details of your instance, and make any necessary changes by clicking the appropriate Edit link.

When you are ready, choose Launch to select a key pair and launch your instance.
4. If the instance fails to launch or the state immediately goes to terminated instead of running, see Troubleshooting Instance Launch Issues (p. 1042).
Launching an AWS Marketplace Instance

You can subscribe to an AWS Marketplace product and launch an instance from the product's AMI using the Amazon EC2 launch wizard. For more information about paid AMIs, see Paid AMIs (p. 60). To cancel your subscription after launch, you first have to terminate all instances running from it. For more information, see Managing Your AWS Marketplace Subscriptions (p. 63).

To launch an instance from the AWS Marketplace using the launch wizard

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the Amazon EC2 dashboard, choose Launch Instance.
3. On the Choose an Amazon Machine Image (AMI) page, choose the AWS Marketplace category on the left. Find a suitable AMI by browsing the categories, or using the search functionality. Choose Select to choose your product.
4. A dialog displays an overview of the product you've selected. You can view the pricing information, as well as any other information that the vendor has provided. When you're ready, choose Continue.
   
   **Note**
   
   You are not charged for using the product until you have launched an instance with the AMI. Take note of the pricing for each supported instance type, as you will be prompted to select an instance type on the next page of the wizard. Additional taxes may also apply to the product.
5. On the Choose an Instance Type page, select the hardware configuration and size of the instance to launch. When you're done, choose Next: Configure Instance Details.
6. On the next pages of the wizard, you can configure your instance, add storage, and add tags. For more information about the different options you can configure, see Launching an Instance Using the Launch Instance Wizard (p. 334). Choose Next until you reach the Configure Security Group page.

   The wizard creates a new security group according to the vendor's specifications for the product. The security group may include rules that allow all IPv4 addresses (0.0.0.0/0) access on SSH (port 22) on Linux or RDP (port 3389) on Windows. We recommend that you adjust these rules to allow only a specific address or range of addresses to access your instance over those ports.

   When you are ready, choose Review and Launch.
7. On the Review Instance Launch page, check the details of the AMI from which you're about to launch the instance, as well as the other configuration details you set up in the wizard. When you're ready, choose Launch to select or create a key pair, and launch your instance.
8. Depending on the product you've subscribed to, the instance may take a few minutes or more to launch. You are first subscribed to the product before your instance can launch. If there are any problems with your credit card details, you will be asked to update your account details. When the launch confirmation page displays, choose View Instances to go to the Instances page.
   
   **Note**
   
   You are charged the subscription price as long as your instance is running, even if it is idle. If your instance is stopped, you may still be charged for storage.
9. When your instance is in the running state, you can connect to it. To do this, select your instance in the list and choose Connect. Follow the instructions in the dialog. For more information about connecting to your instance, see Connecting to Your Windows Instance (p. 379).
   
   **Important**
   
   Check the vendor's usage instructions carefully, as you may need to use a specific user name to log in to the instance. For more information about accessing your subscription details, see Managing Your AWS Marketplace Subscriptions (p. 63).
10. If the instance fails to launch or the state immediately goes to terminated instead of running, see Troubleshooting Instance Launch Issues (p. 1042).
Launching an AWS Marketplace AMI Instance Using the API and CLI

To launch instances from AWS Marketplace products using the API or command line tools, first ensure that you are subscribed to the product. You can then launch an instance with the product’s AMI ID using the following methods:

<table>
<thead>
<tr>
<th>Method</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS CLI</td>
<td>Use the <code>run-instances</code> command, or see the following topic for more information: Launching an Instance.</td>
</tr>
<tr>
<td>AWS Tools for Windows PowerShell</td>
<td>Use the <code>New-EC2Instance</code> command, or see the following topic for more information: Launch an Amazon EC2 Instance Using Windows PowerShell</td>
</tr>
<tr>
<td>Query API</td>
<td>Use the <code>RunInstances</code> request.</td>
</tr>
</tbody>
</table>

Launching an EC2 Fleet

An EC2 Fleet contains the configuration information to launch a fleet—or group—of instances. In a single API call, a fleet can launch multiple instance types across multiple Availability Zones, using the On-Demand Instance, Reserved Instance, and Spot Instance purchasing options together. Using EC2 Fleet, you can:

- Define separate On-Demand and Spot capacity targets and the maximum amount you’re willing to pay per hour
- Specify the instance types that work best for your applications
- Specify how Amazon EC2 should distribute your fleet capacity within each purchasing option

You can also set a maximum amount per hour that you're willing to pay for your fleet, and EC2 Fleet launches instances until it reaches the maximum amount. When the maximum amount you're willing to pay is reached, the fleet stops launching instances even if it hasn't met the target capacity.

The EC2 Fleet attempts to launch the number of instances that are required to meet the target capacity specified in your request. If you specified a total maximum price per hour, it fulfills the capacity until it reaches the maximum amount that you’re willing to pay. The fleet can also attempt to maintain its target Spot capacity if your Spot Instances are interrupted. For more information, see How Spot Instances Work (p. 234).

You can specify an unlimited number of instance types per EC2 Fleet. Those instance types can be provisioned using both On-Demand and Spot purchasing options. You can also specify multiple Availability Zones, specify different maximum Spot prices for each instance, and choose additional
Spot options for each fleet. Amazon EC2 uses the specified options to provision capacity when the fleet launches.

While the fleet is running, if Amazon EC2 reclaims a Spot Instance because of a price increase or instance failure, EC2 Fleet can try to replace the instances with any of the instance types that you specify. This makes it easier to regain capacity during a spike in Spot pricing. You can develop a flexible and elastic resourcing strategy for each fleet. For example, within specific fleets, your primary capacity can be On-Demand supplemented with less-expensive Spot capacity if available.

If you have Reserved Instances and you specify On-Demand Instances in your fleet, EC2 Fleet uses your Reserved Instances. For example, if your fleet specifies an On-Demand Instance as `c4.large`, and you have Reserved Instances for `c4.large`, you receive the Reserved Instance pricing.

There is no additional charge for using EC2 Fleet. You pay only for the EC2 instances that the fleet launches for you.

**Contents**

- EC2 Fleet Limitations (p. 352)
- EC2 Fleet Limits (p. 352)
- EC2 Fleet Configuration Strategies (p. 353)
- Managing an EC2 Fleet (p. 362)

**EC2 Fleet Limitations**

The following limitations apply to EC2 Fleet:

- EC2 Fleet is available only through the API or AWS CLI.
- An EC2 Fleet request can't span AWS Regions. You need to create a separate EC2 Fleet for each Region.
- An EC2 Fleet request can't span different subnets from the same Availability Zone.

**EC2 Fleet Limits**

The usual Amazon EC2 limits apply to instances launched by an EC2 Fleet, such as Spot request price limits, instance limits, and volume limits. In addition, the following limits apply:

- The number of active EC2 Fleets per AWS Region: 1,000 *
- The number of launch specifications per fleet: 50 †
- The size of the user data in a launch specification: 16 KB †
- The target capacity per EC2 Fleet: 10,000
- The target capacity across all EC2 Fleets in a Region: 100,000 *

If you need more than the default limits for target capacity, complete the AWS Support Center Create case form to request a limit increase. For Limit type, choose EC2 Fleet, choose a Region, and then choose Target Fleet Capacity per Fleet (in units) or Target Fleet Capacity per Region (in units), or both.

* These limits apply to both your EC2 Fleets and your Spot Fleets.
† These are hard limits. You cannot request a limit increase for these limits.

**T3 Instances**

If you plan to use your T3 Spot Instances immediately and for a short duration, with no idle time for accruing CPU credits, we recommend that you launch your T3 Spot Instances in standard (p. 142) mode to avoid paying higher costs.
If you launch your T3 Spot Instances in unlimited (p. 135) mode and burst CPU immediately, you'll spend surplus credits for bursting. If you use the instance for a short duration, your instance doesn't have time to accrue CPU credits to pay down the surplus credits, and you are charged for the surplus credits when you terminate your instance.

Unlimited mode for T3 Spot Instances is suitable only if the instance runs for long enough to accrue CPU credits for bursting. Otherwise, paying for surplus credits makes T3 Spot Instances more expensive than M5 or C5 instances.

T2 Instances

Launch credits are meant to provide a productive initial launch experience for T2 instances by providing sufficient compute resources to configure the instance. Repeated launches of T2 instances to access new launch credits is not permitted. If you require sustained CPU, you can earn credits (by idling over some period), use T2 Unlimited (p. 135), or use an instance type with dedicated CPU (for example, c4.large).

EC2 Fleet Configuration Strategies

An EC2 Fleet is a group of On-Demand Instances and Spot Instances. The EC2 Fleet attempts to launch the number of instances that are required to meet the target capacity that you specify in the fleet request. The fleet can comprise only On-Demand Instances, only Spot Instances, or a combination of both On-Demand Instances and Spot Instances. The request for Spot Instances is fulfilled if there is available capacity and the maximum price per hour for your request exceeds the Spot price. The fleet also attempts to maintain its target capacity if your Spot Instances are interrupted.

You can also set a maximum amount per hour that you're willing to pay for your fleet, and EC2 Fleet launches instances until it reaches the maximum amount. When the maximum amount you're willing to pay is reached, the fleet stops launching instances even if it hasn't met the target capacity.

A Spot Instance pool is a set of unused EC2 instances with the same instance type, operating system, Availability Zone, and network platform. When you create an EC2 Fleet, you can include multiple launch specifications, which vary by instance type, Availability Zone, subnet, and maximum price. The fleet selects the Spot Instance pools that are used to fulfill the request, based on the launch specifications included in your request, and the configuration of the request. The Spot Instances come from the selected pools.

An EC2 Fleet enables you to provision large amounts of EC2 capacity that makes sense for your application based on number of cores or instances, or amount of memory. For example, you can specify an EC2 Fleet to launch a target capacity of 200 instances, of which 150 are On-Demand Instances and the rest are Spot Instances. Or you can request 1000 cores with a minimum of 2 GB of RAM per core. The fleet determines the combination of Amazon EC2 options to launch that capacity at the absolute lowest cost.

Use the appropriate configuration strategies to create an EC2 Fleet that meets your needs.

Contents
- Planning an EC2 Fleet (p. 354)
- EC2 Fleet Request Types (p. 354)
- Allocation Strategies for Spot Instances (p. 355)
- Configuring EC2 Fleet for On-Demand Backup (p. 356)
- Maximum Price Overrides (p. 356)
- Control Spending (p. 357)
- EC2 Fleet Instance Weighting (p. 357)
- Tutorial: Using EC2 Fleet with Instance Weighting (p. 359)
Planning an EC2 Fleet

When planning your EC2 Fleet, we recommend that you do the following:

- Determine whether you want to create an EC2 Fleet that submits a synchronous or asynchronous one-time request for the desired target capacity, or one that maintains a target capacity over time. For more information, see EC2 Fleet Request Types (p. 354).
- Determine the instance types that meet your application requirements.
- If you plan to include Spot Instances in your EC2 Fleet, review Spot Best Practices before you create the fleet. Use these best practices when you plan your fleet so that you can provision the instances at the lowest possible price.
- Determine the target capacity for your EC2 Fleet. You can set target capacity in instances or in custom units. For more information, see EC2 Fleet Instance Weighting (p. 357).
- Determine what portion of the EC2 Fleet target capacity must be On-Demand capacity and Spot capacity. You can specify 0 for On-Demand capacity or Spot capacity, or both.
- Determine your price per unit, if you are using instance weighting. To calculate the price per unit, divide the price per instance hour by the number of units (or weight) that this instance represents. If you are not using instance weighting, the default price per unit is the price per instance hour.
- Determine the maximum amount per hour that you’re willing to pay for your fleet. For more information, see Control Spending (p. 357).
- Review the possible options for your EC2 Fleet. For more information, see the EC2 Fleet JSON Configuration File Reference (p. 367). For EC2 Fleet configuration examples, see EC2 Fleet Example Configurations (p. 376).

EC2 Fleet Request Types

There are three types of EC2 Fleet requests:

- instant

  If you configure the request type as instant, EC2 Fleet places a synchronous one-time request for your desired capacity. In the API response, it returns the instances that launched, along with errors for those instances that could not be launched.

- request

  If you configure the request type as request, EC2 Fleet places an asynchronous one-time request for your desired capacity. Thereafter, if capacity is diminished because of Spot interruptions, the fleet does not attempt to replenish Spot Instances, nor does it submit requests in alternative Spot Instance pools if capacity is unavailable.

- maintain

  (Default) If you configure the request type as maintain, EC2 Fleet places an asynchronous request for your desired capacity, and maintains capacity by automatically replenishing any interrupted Spot Instances.

You cannot modify the target capacity of an instant or request EC2 Fleet request after it is submitted. To change the target capacity of an instant or request fleet request, delete the fleet and create a new one.

All three types of requests benefit from an allocation strategy. For more information, see Allocation Strategies for Spot Instances (p. 355).
Allocation Strategies for Spot Instances

The allocation strategy for your EC2 Fleet determines how it fulfills your request for Spot Instances from the possible Spot instance pools represented by its launch specifications. The following are the allocation strategies that you can specify in your fleet:

lowestPrice

The Spot Instances come from the pool with the lowest price. This is the default strategy.

diversified

The Spot Instances are distributed across all pools.

capacityOptimized

The Spot Instances come from the pool with optimal capacity for the number of instances that are launching.

InstancePoolsToUseCount

The Spot Instances are distributed across the number of Spot pools that you specify. This parameter is valid only when used in combination with lowestPrice.

Maintaining Target Capacity

After Spot Instances are terminated due to a change in the Spot price or available capacity of a Spot Instance pool, an EC2 Fleet of type maintain launches replacement Spot Instances. If the allocation strategy is lowestPrice, the fleet launches replacement instances in the pool where the Spot price is currently the lowest. If the allocation strategy is lowestPrice in combination with InstancePoolsToUseCount, the fleet selects the Spot pools with the lowest price and launches Spot Instances across the number of Spot pools that you specify. If the allocation strategy is capacityOptimized, the fleet launches replacement instances in the pool that has the most available Spot Instance capacity. If the allocation strategy is diversified, the fleet distributes the replacement Spot Instances across the remaining pools.

Configuring EC2 Fleet for Cost Optimization

To optimize the costs for your use of Spot Instances, specify the lowestPrice allocation strategy so that EC2 Fleet automatically deploys the least expensive combination of instance types and Availability Zones based on the current Spot price.

For On-Demand Instance target capacity, EC2 Fleet always selects the cheapest instance type based on the public On-Demand price, while continuing to follow the allocation strategy (either lowestPrice, capacityOptimized, or diversified) for Spot Instances.

Configuring EC2 Fleet for Cost Optimization and Diversification

To create a fleet of Spot Instances that is both cheap and diversified, use the lowestPrice allocation strategy in combination with InstancePoolsToUseCount. EC2 Fleet automatically deploys the least expensive combination of instance types and Availability Zones based on the current Spot price across the number of Spot pools that you specify. This combination can be used to avoid the most expensive Spot Instances.

Configuring EC2 Fleet for Capacity Optimization

With Spot Instances, pricing changes slowly over time based on long-term trends in supply and demand, but capacity fluctuates in real time. The capacityOptimized strategy automatically launches Spot Instances into the most available pools by looking at real-time capacity data and predicting which are the most available. This works well for workloads such as big data and analytics, image and media rendering, machine learning, and high performance computing that may have a higher cost of
choosing the appropriate allocation strategy

you can optimize your fleet based on your use case.

if your fleet is small or runs for a short time, the probability that your spot instances will be interrupted is low, even with all of the instances in a single spot instance pool. therefore, the lowestPrice strategy is likely to meet your needs while providing the lowest cost.

if your fleet is large or runs for a long time, you can improve the availability of your fleet by distributing the spot instances across multiple pools. for example, if your EC2 fleet specifies 10 pools and a target capacity of 100 instances, the fleet launches 10 spot instances in each pool. if the spot price for one pool exceeds your maximum price for this pool, only 10% of your fleet is affected. using this strategy also makes your fleet less sensitive to increases in the spot price in any one pool over time.

with the diversified strategy, the EC2 fleet does not launch spot instances into any pools with a spot price that is equal to or higher than the On-Demand price.

To create a cheap and diversified fleet, use the lowestPrice strategy in combination with InstancePoolsToUseCount. You can use a low or high number of spot pools across which to allocate your spot instances. For example, if you run batch processing, we recommend specifying a low number of spot pools (for example, InstancePoolsToUseCount=2) to ensure that your queue always has compute capacity while maximizing savings. If you run a web service, we recommend specifying a high number of spot pools (for example, InstancePoolsToUseCount=10) to minimize the impact if a spot instance pool becomes temporarily unavailable.

if your fleet runs workloads that may have a higher cost of interruption associated with restarting work and checkpointing, then use the capacityOptimized strategy. this strategy offers the possibility of fewer interruptions, which can lower the overall cost of your workload.

configuring EC2 fleet for on-demand backup

if you have urgent, unpredictable scaling needs, such as a news website that must scale during a major news event or game launch, we recommend that you specify alternative instance types for your on-demand instances, in the event that your preferred option does not have sufficient available capacity. For example, you might prefer c5.2xlarge on-demand instances, but if there is insufficient available capacity, you’d be willing to use some c4.2xlarge instances during peak load. In this case, EC2 fleet attempts to fulfill all of your target capacity using c5.2xlarge instances, but if there is insufficient capacity, it automatically launches c4.2xlarge instances to fulfill the target capacity.

prioritizing instance types for on-demand capacity

when EC2 fleet attempts to fulfill your on-demand capacity, it defaults to launching the lowest-priced instance type first. if AllocationStrategy is set to prioritized, EC2 fleet uses priority to determine which instance type to use first in fulfilling on-demand capacity. the priority is assigned to the launch template override, and the highest priority is launched first.

for example, you have configured three launch template overrides, each with a different instance type: c3.large, c4.large, and c5.large. the on-demand price for c5.large is less than the price for c4.large. c3.large is the cheapest. if you do not use priority to determine the order, the fleet fulfills on-demand capacity by starting with c3.large, and then c5.large. because you often have unused reserved instances for c4.large, you can set the launch template override priority so that the order is c4.large, c3.large, and then c5.large.

maximum price overrides

Each EC2 fleet can either include a global maximum price, or use the default (the on-demand price). The fleet uses this as the default maximum price for each of its launch specifications.
You can optionally specify a maximum price in one or more launch specifications. This price is specific to the launch specification. If a launch specification includes a specific price, the EC2 Fleet uses this maximum price, overriding the global maximum price. Any other launch specifications that do not include a specific maximum price still use the global maximum price.

Control Spending

EC2 Fleet stops launching instances when it has met one of the following parameters: the TotalTargetCapacity or the MaxTotalPrice (the maximum amount you’re willing to pay). To control the amount you pay per hour for your fleet, you can specify the MaxTotalPrice. When the maximum total price is reached, EC2 Fleet stops launching instances even if it hasn’t met the target capacity.

The following examples show two different scenarios. In the first, EC2 Fleet stops launching instances when it has met the target capacity. In the second, EC2 Fleet stops launching instances when it has reached the maximum amount you’re willing to pay (MaxTotalPrice).

Example: Stop launching instances when target capacity is reached

Given a request for m4.large On-Demand Instances, where:

- On-Demand Price: $0.10 per hour
- OnDemandTargetCapacity: 10
- MaxTotalPrice: $1.50

EC2 Fleet launches 10 On-Demand Instances because the total of $1.00 (10 instances x $0.10) does not exceed the MaxTotalPrice of $1.50 for On-Demand Instances.

Example: Stop launching instances when maximum total price is reached

Given a request for m4.large On-Demand Instances, where:

- On-Demand Price: $0.10 per hour
- OnDemandTargetCapacity: 10
- MaxTotalPrice: $0.80

If EC2 Fleet launches the On-Demand target capacity (10 On-Demand Instances), the total cost per hour would be $1.00. This is more than the amount ($0.80) specified for MaxTotalPrice for On-Demand Instances. To prevent spending more than you’re willing to pay, EC2 Fleet launches only 8 On-Demand Instances (below the On-Demand target capacity) because launching more would exceed the MaxTotalPrice for On-Demand Instances.

EC2 Fleet Instance Weighting

When you create an EC2 Fleet, you can define the capacity units that each instance type would contribute to your application’s performance. You can then adjust your maximum price for each launch specification by using instance weighting.

By default, the price that you specify is per instance hour. When you use the instance weighting feature, the price that you specify is per unit hour. You can calculate your price per unit hour by dividing your price for an instance type by the number of units that it represents. EC2 Fleet calculates the number of instances to launch by dividing the target capacity by the instance weight. If the result isn't an integer, the fleet rounds it up to the next integer, so that the size of your fleet is not below its target capacity. The fleet can select any pool that you specify in your launch specification, even if the capacity of the instances launched exceeds the requested target capacity.
The following table includes examples of calculations to determine the price per unit for an EC2 Fleet with a target capacity of 10.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Instance weight</th>
<th>Target capacity</th>
<th>Number of instances launched</th>
<th>Price per instance hour</th>
<th>Price per unit hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.xlarge</td>
<td>2</td>
<td>10</td>
<td>5</td>
<td>$0.05</td>
<td>$0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10 divided by 2)</td>
<td></td>
<td>(.05 divided by 2)</td>
</tr>
<tr>
<td>r3.8xlarge</td>
<td>8</td>
<td>10</td>
<td>2</td>
<td>$0.10</td>
<td>$0.0125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(10 divided by 8, result rounded up)</td>
<td></td>
<td>(.10 divided by 8)</td>
</tr>
</tbody>
</table>

Use EC2 Fleet instance weighting as follows to provision the target capacity that you want in the pools with the lowest price per unit at the time of fulfillment:

1. Set the target capacity for your EC2 Fleet either in instances (the default) or in the units of your choice, such as virtual CPUs, memory, storage, or throughput.
2. Set the price per unit.
3. For each launch specification, specify the weight, which is the number of units that the instance type represents toward the target capacity.

**Instance Weighting Example**

Consider an EC2 Fleet request with the following configuration:

- A target capacity of 24
- A launch specification with an instance type r3.2xlarge and a weight of 6
- A launch specification with an instance type c3.xlarge and a weight of 5

The weights represent the number of units that instance type represents toward the target capacity. If the first launch specification provides the lowest price per unit (price for r3.2xlarge per instance hour divided by 6), the EC2 Fleet would launch four of these instances (24 divided by 6).

If the second launch specification provides the lowest price per unit (price for c3.xlarge per instance hour divided by 5), the EC2 Fleet would launch five of these instances (24 divided by 5, result rounded up).

**Instance Weighting and Allocation Strategy**

Consider an EC2 Fleet request with the following configuration:

- A target capacity of 30 Spot Instances
- A launch specification with an instance type c3.2xlarge and a weight of 8
- A launch specification with an instance type m3.xlarge and a weight of 8
- A launch specification with an instance type r3.xlarge and a weight of 8

The EC2 Fleet would launch four instances (30 divided by 8, result rounded up). With the lowestPrice strategy, all four instances come from the pool that provides the lowest price per unit. With the
diversified strategy, the fleet launches one instance in each of the three pools, and the fourth instance in whichever of the three pools provides the lowest price per unit.

**Tutorial: Using EC2 Fleet with Instance Weighting**

This tutorial uses a fictitious company called Example Corp to illustrate the process of requesting an EC2 Fleet using instance weighting.

**Objective**

Example Corp, a pharmaceutical company, wants to use the computational power of Amazon EC2 for screening chemical compounds that might be used to fight cancer.

**Planning**

Example Corp first reviews Spot Best Practices. Next, Example Corp determines the requirements for their EC2 Fleet.

**Instance Types**

Example Corp has a compute- and memory-intensive application that performs best with at least 60 GB of memory and eight virtual CPUs (vCPUs). They want to maximize these resources for the application at the lowest possible price. Example Corp decides that any of the following EC2 instance types would meet their needs:

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Memory (GiB)</th>
<th>vCPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.2xlarge</td>
<td>61</td>
<td>8</td>
</tr>
<tr>
<td>r3.4xlarge</td>
<td>122</td>
<td>16</td>
</tr>
<tr>
<td>r3.8xlarge</td>
<td>244</td>
<td>32</td>
</tr>
</tbody>
</table>

**Target Capacity in Units**

With instance weighting, target capacity can equal a number of instances (the default) or a combination of factors such as cores (vCPUs), memory (GiBs), and storage (GBs). By considering the base for their application (60 GB of RAM and eight vCPUs) as one unit, Example Corp decides that 20 times this amount would meet their needs. So the company sets the target capacity of their EC2 Fleet request to 20.

**Instance Weights**

After determining the target capacity, Example Corp calculates instance weights. To calculate the instance weight for each instance type, they determine the units of each instance type that are required to reach the target capacity as follows:

- r3.2xlarge (61.0 GB, 8 vCPUs) = 1 unit of 20
- r3.4xlarge (122.0 GB, 16 vCPUs) = 2 units of 20
- r3.8xlarge (244.0 GB, 32 vCPUs) = 4 units of 20

Therefore, Example Corp assigns instance weights of 1, 2, and 4 to the respective launch configurations in their EC2 Fleet request.

**Price Per Unit Hour**
Example Corp uses the On-Demand price per instance hour as a starting point for their price. They could also use recent Spot prices, or a combination of the two. To calculate the price per unit hour, they divide their starting price per instance hour by the weight. For example:

<table>
<thead>
<tr>
<th>Instance type</th>
<th>On-Demand price</th>
<th>Instance weight</th>
<th>Price per unit hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.2xLarge</td>
<td>$0.7</td>
<td>1</td>
<td>$0.7</td>
</tr>
<tr>
<td>r3.4xLarge</td>
<td>$1.4</td>
<td>2</td>
<td>$0.7</td>
</tr>
<tr>
<td>r3.8xLarge</td>
<td>$2.8</td>
<td>4</td>
<td>$0.7</td>
</tr>
</tbody>
</table>

Example Corp could use a global price per unit hour of $0.7 and be competitive for all three instance types. They could also use a global price per unit hour of $0.7 and a specific price per unit hour of $0.9 in the r3.8xlarge launch specification.

**Verifying Permissions**

Before creating an EC2 Fleet, Example Corp verifies that it has an IAM role with the required permissions. For more information, see EC2 Fleet Prerequisites (p. 363).

**Creating the EC2 Fleet**

Example Corp creates a file, config.json, with the following configuration for its EC2 Fleet.

```json
{
   "LaunchTemplateConfigs": [
       {
           "LaunchTemplateSpecification": {
               "LaunchTemplateId": "lt-07b3bc7625cdab851",
               "Version": "1"
           },
           "Overrides": [
               {
                   "InstanceType": "r3.2xlarge",
                   "SubnetId": "subnet-482e4972",
                   "WeightedCapacity": 1
               },
               {
                   "InstanceType": "r3.4xlarge",
                   "SubnetId": "subnet-482e4972",
                   "WeightedCapacity": 2
               },
               {
                   "InstanceType": "r3.8xlarge",
                   "MaxPrice": "0.90",
                   "SubnetId": "subnet-482e4972",
                   "WeightedCapacity": 4
               }
           ]
       },
       "TargetCapacitySpecification": {
           "TotalTargetCapacity": 20,
           "DefaultTargetCapacityType": "spot"
       }
   ]
}
```

Example Corp creates the EC2 Fleet using the following `create-fleet` command.
Fulfillment

The allocation strategy determines which Spot Instance pools your Spot Instances come from.

With the `lowestPrice` strategy (which is the default strategy), the Spot Instances come from the pool with the lowest price per unit at the time of fulfillment. To provide 20 units of capacity, the EC2 Fleet launches either 20 `r3.2xlarge` instances (20 divided by 1), 10 `r3.4xlarge` instances (20 divided by 2), or 5 `r3.8xlarge` instances (20 divided by 4).

If Example Corp used the `diversified` strategy, the Spot Instances would come from all three pools. The EC2 Fleet would launch 6 `r3.2xlarge` instances (which provide 6 units), 3 `r3.4xlarge` instances (which provide 6 units), and 2 `r3.8xlarge` instances (which provide 8 units), for a total of 20 units.

Tutorial: Using EC2 Fleet with On-Demand as the Primary Capacity

This tutorial uses a fictitious company called ABC Online to illustrate the process of requesting an EC2 Fleet with On-Demand as the primary capacity, and Spot capacity if available.

Objective

ABC Online, a restaurant delivery company, wants to be able to provision Amazon EC2 capacity across EC2 instance types and purchasing options to achieve their desired scale, performance, and cost.

Planning

ABC Online requires a fixed capacity to operate during peak periods, but would like to benefit from increased capacity at a lower price. ABC Online determines the following requirements for their EC2 Fleet:

- **On-Demand Instance capacity** – ABC Online requires 15 On-Demand Instances to ensure that they can accommodate traffic at peak periods.
- **Spot Instance capacity** – ABC Online would like to improve performance, but at a lower price, by provisioning 5 Spot Instances.

Verifying Permissions

Before creating an EC2 Fleet, ABC Online verifies that it has an IAM role with the required permissions. For more information, see EC2 Fleet Prerequisites (p. 363).

Creating the EC2 Fleet

ABC Online creates a file, `config.json`, with the following configuration for its EC2 Fleet.

```json
{
   "LaunchTemplateConfigs": [
      {
         "LaunchTemplateSpecification": {
            "LaunchTemplateId": "lt-07b3bc7625cdab851",
            "Version": "2"
         }
      }
   ],
   "TargetCapacitySpecification": {
```

```
ABC Online creates the EC2 Fleet using the following create-fleet command.

```bash
aws ec2 create-fleet --cli-input-json file://config.json
```

For more information, see Creating an EC2 Fleet (p. 370).

**Fulfillment**

The allocation strategy determines that the On-Demand capacity is always fulfilled, while the balance of the target capacity is fulfilled as Spot if there is capacity and availability.

**Managing an EC2 Fleet**

To use an EC2 Fleet, you create a request that includes the total target capacity, On-Demand capacity, Spot capacity, one or more launch specifications for the instances, and the maximum price that you are willing to pay. The fleet request must include a launch template that defines the information that the fleet needs to launch an instance, such as an AMI, instance type, subnet or Availability Zone, and one or more security groups. You can specify launch specification overrides for the instance type, subnet, Availability Zone, and maximum price you're willing to pay, and you can assign weighted capacity to each launch specification override.

If your fleet includes Spot Instances, Amazon EC2 can attempt to maintain your fleet target capacity as Spot prices change.

An EC2 Fleet request remains active until it expires or you delete it. When you delete a fleet, you can specify whether deletion terminates the instances in that fleet.

**Contents**

- EC2 Fleet Request States (p. 362)
- EC2 Fleet Prerequisites (p. 363)
- EC2 Fleet Health Checks (p. 365)
- Generating an EC2 Fleet JSON Configuration File (p. 366)
- Creating an EC2 Fleet (p. 370)
- Tagging an EC2 Fleet (p. 373)
- Monitoring Your EC2 Fleet (p. 362)
- Modifying an EC2 Fleet (p. 375)
- Deleting an EC2 Fleet (p. 375)
- EC2 Fleet Example Configurations (p. 376)

**EC2 Fleet Request States**

An EC2 Fleet request can be in one of the following states:

- **submitted** – The EC2 Fleet request is being evaluated and Amazon EC2 is preparing to launch the target number of instances, which can include On-Demand Instances, Spot Instances, or both.
- **active** – The EC2 Fleet request has been validated and Amazon EC2 is attempting to maintain the target number of running instances. The request remains in this state until it is modified or deleted.
• **modifying** – The EC2 Fleet request is being modified. The request remains in this state until the modification is fully processed or the request is deleted. Only a maintain request type can be modified. This state does not apply to other request types.

• **deleted_running** – The EC2 Fleet request is deleted and does not launch additional instances. Its existing instances continue to run until they are interrupted or terminated. The request remains in this state until all instances are interrupted or terminated.

• **deleted_terminating** – The EC2 Fleet request is deleted and its instances are terminating. The request remains in this state until all instances are terminated.

• **deleted** – The EC2 Fleet is deleted and has no running instances. The request is deleted two days after its instances are terminated.

The following illustration represents the transitions between the EC2 Fleet request states. If you exceed your fleet limits, the request is deleted immediately.

---

**EC2 Fleet Prerequisites**

To create an EC2 Fleet, the following prerequisites must be in place.

**Launch Template**

A launch template includes information about the instances to launch, such as the instance type, Availability Zone, and the maximum price that you are willing to pay. For more information, see Launching an Instance from a Launch Template (p. 338).

**Service-Linked Role for EC2 Fleet**

The AWSServiceRoleForEC2Fleet role grants the EC2 Fleet permission to request, launch, terminate, and tag instances on your behalf. Amazon EC2 uses this service-linked role to complete the following actions:

- `ec2:RequestSpotInstances` – Request Spot Instances.
- `ec2:TerminateInstances` – Terminate Spot Instances.
- `ec2:DescribeImages` – Describe Amazon Machine Images (AMI) for the Spot Instances.
- `ec2:DescribeInstanceStatus` – Describe the status of the Spot Instances.
- `ec2:DescribeSubnets` – Describe the subnets for Spot Instances.
• `ec2:CreateTags` – Add system tags to Spot Instances.

Ensure that this role exists before you use the AWS CLI or an API to create an EC2 Fleet. To create the role, use the IAM console as follows.

**To create the AWSServiceRoleForEC2Fleet role for EC2 Fleet**

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose **Roles**, and then choose **Create role**.
3. For **Select type of trusted entity**, choose **AWS service**.
4. For **Choose the service that will use this role**, choose **EC2 - Fleet**, and then choose Next: Permissions, Next: Tags, and Next: Review.
5. On the Review page, choose Create role.

If you no longer need to use EC2 Fleet, we recommend that you delete the `AWSServiceRoleForEC2Fleet` role. After this role is deleted from your account, you can create the role again if you create another fleet.

For more information, see Using Service-Linked Roles in the *IAM User Guide*.

**Granting Access to CMKs for Use with Encrypted AMIs and EBS Snapshots**

If you specify an encrypted AMI (p. 67) or an encrypted Amazon EBS snapshot (p. 865) in your EC2 Fleet and you use a customer managed customer master key (CMK) for encryption, you must grant the `AWSServiceRoleForEC2Fleet` role permission to use the CMK so that Amazon EC2 can launch instances on your behalf. To do this, you must add a grant to the CMK, as shown in the following procedure.

When providing permissions, grants are an alternative to key policies. For more information, see Using Grants and Using Key Policies in AWS KMS in the *AWS Key Management Service Developer Guide*.

**To grant the AWSServiceRoleForEC2Fleet role permissions to use the CMK**

• Use the `create-grant` command to add a grant to the CMK and to specify the principal (the `AWSServiceRoleForEC2Fleet` service-linked role) that is given permission to perform the operations that the grant permits. The CMK is specified by the `key-id` parameter and the ARN of the CMK. The principal is specified by the `grantee-principal` parameter and the ARN of the `AWSServiceRoleForEC2Fleet` service-linked role.

The following example is formatted for legibility.

```bash
aws kms create-grant
  --region us-east-1
  --key-id arn:aws:kms:us-east-1:123456789012:alias/encryption-key
  --grantee-principal arn:aws:iam::111122223333:role/EC2Fleet
  --operations "Decrypt" "Encrypt" "GenerateDataKey" "GenerateDataKeyWithoutPlaintext" "CreateGrant" "DescribeKey" "ReEncryptFrom" "ReEncryptTo"
```

**EC2 Fleet and IAM Users**

If your IAM users will create or manage an EC2 Fleet, be sure to grant them the required permissions as follows.

**To grant an IAM user permissions for EC2 Fleet**

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose **Policies**.
3. Choose **Create policy**.
4. On the **Create policy** page, choose the **JSON** tab, replace the text with the following, and choose **Review policy**.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": ["ec2:*"],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": ["iam:ListRoles", "iam:PassRole", "iam:ListInstanceProfiles"],
            "Resource": "*"
        }
    ]
}
```

The `ec2:*` grants an IAM user permission to call all Amazon EC2 API actions. To limit the user to specific Amazon EC2 API actions, specify those actions instead.

An IAM user must have permission to call the `iam:ListRoles` action to enumerate existing IAM roles, the `iam:PassRole` action to specify the EC2 Fleet role, and the `iam:ListInstanceProfiles` action to enumerate existing instance profiles.

(Optional) To enable an IAM user to create roles or instance profiles using the IAM console, you must also add the following actions to the policy:

- `iam:AddRoleToInstanceProfile`
- `iam:AttachRolePolicy`
- `iam:CreateInstanceProfile`
- `iam:CreateRole`
- `iam:GetRole`
- `iam:ListPolicies`

5. On the **Review policy** page, enter a policy name and description, and choose **Create policy**.

6. In the navigation pane, choose **Users** and select the user.

7. On the **Permissions** tab, choose **Add permissions**.

8. Choose **Attach existing policies directly**. Select the policy that you created earlier and choose **Next: Review**.

9. Choose **Add permissions**.

**EC2 Fleet Health Checks**

EC2 Fleet checks the health status of the instances in the fleet every two minutes. The health status of an instance is either healthy or unhealthy. The fleet determines the health status of an instance using the status checks provided by Amazon EC2. If the status of either the instance status check or the system status check is impaired for three consecutive health checks, the health status of the instance is unhealthy. Otherwise, the health status is healthy. For more information, see **Status Checks for Your Instances** (p. 535).
You can configure your EC2 Fleet to replace unhealthy instances. After enabling health check replacement, an instance is replaced after its health status is reported as unhealthy. The fleet could go below its target capacity for up to a few minutes while an unhealthy instance is being replaced.

Requirements

- Health check replacement is supported only with EC2 Fleets that maintain a target capacity, not with one-time fleets.
- You can configure your EC2 Fleet to replace unhealthy instances only when you create it.
- IAM users can use health check replacement only if they have permission to call the ec2:DescribeInstanceStatus action.

Generating an EC2 Fleet JSON Configuration File

To create an EC2 Fleet, you need only specify the launch template, total target capacity, and whether the default purchasing option is On-Demand or Spot. If you do not specify a parameter, the fleet uses the default value. To view the full list of fleet configuration parameters, you can generate a JSON file as follows.

To generate a JSON file with all possible EC2 Fleet parameters using the command line

- Use the create-fleet (AWS CLI) command and the --generate-cli-skeleton parameter to generate an EC2 Fleet JSON file:

  ```
  aws ec2 create-fleet --generate-cli-skeleton
  ```

The following EC2 Fleet parameters are available:

```json
{
  "DryRun": true,
  "ClientToken": "",
  "SpotOptions": {
    "AllocationStrategy": "lowestPrice",
    "InstanceInterruptionBehavior": "hibernate",
    "InstancePoolsToUseCount": 0
  },
  "OnDemandOptions": {
    "AllocationStrategy": "prioritized",
    "SingleInstanceType": true,
    "SingleAvailabilityZone": true,
    "MaxTotalPrice": 0,
    "MinTargetCapacity": 0
  },
  "ExcessCapacityTerminationPolicy": "termination",
  "LaunchTemplateConfigs": [
    {
      "LaunchTemplateSpecification": {
        "LaunchTemplateId": "",
        "LaunchTemplateName": "",
        "Version": ""
      },
      "Overrides": [
        {
          "InstanceType": "t2.micro",
          "MaxPrice": "",
          "SubnetId": ""
        }
      ]
    }
  ]
}
```
EC2 Fleet JSON Configuration File Reference

**Note**
Use lowercase for all parameter values; otherwise, you get an error when Amazon EC2 uses the JSON file to launch the EC2 Fleet.

**AllocationStrategy (for SpotOptions)**
(Optional) Indicates how to allocate the Spot Instance target capacity across the Spot Instance pools specified by the EC2 Fleet. Valid values are `lowestPrice` and `diversified`. The default is `lowestPrice`. Specify the allocation strategy that meets your needs. For more information, see Allocation Strategies for Spot Instances (p. 355).

**InstanceInterruptionBehavior**
(Optional) The behavior when a Spot Instance is interrupted. Valid values are `hibernate`, `stop`, and `terminate`. By default, the Spot service terminates Spot Instances when they are interrupted. If the fleet type is `maintain`, you can specify that the Spot service hibernates or stops Spot Instances when they are interrupted.

**InstancePoolsToUseCount**
The number of Spot pools across which to allocate your target Spot capacity. Valid only when Spot AllocationStrategy is set to `lowestPrice`. EC2 Fleet selects the cheapest Spot pools and evenly allocates your target Spot capacity across the number of Spot pools that you specify.
**SingleInstanceType**

Indicates that the fleet uses a single instance type to launch all Spot Instances in the fleet.

**SingleAvailabilityZone**

Indicates that the fleet launches all Spot Instances into a single Availability Zone.

**MaxTotalPrice**

The maximum amount per hour for Spot Instances that you're willing to pay.

**MinTargetCapacity**

The minimum target capacity for Spot Instances in the fleet. If the minimum target capacity is not reached, the fleet launches no instances.

**AllocationStrategy (for OnDemandOptions)**

The order of the launch template overrides to use in fulfilling On-Demand capacity. If you specify `lowestPrice`, EC2 Fleet uses price to determine the order, launching the lowest price first. If you specify `prioritized`, EC2 Fleet uses the priority that you assigned to each launch template override, launching the highest priority first. If you do not specify a value, EC2 Fleet defaults to `lowestPrice`.

**SingleInstanceType**

Indicates that the fleet uses a single instance type to launch all On-Demand Instances in the fleet.

**SingleAvailabilityZone**

Indicates that the fleet launches all On-Demand Instances into a single Availability Zone.

**MaxTotalPrice**

The maximum amount per hour for On-Demand Instances that you're willing to pay.

**MinTargetCapacity**

The minimum target capacity for On-Demand Instances in the fleet. If the minimum target capacity is not reached, the fleet launches no instances.

**ExcessCapacityTerminationPolicy**

(Optional) Indicates whether running instances should be terminated if the total target capacity of the EC2 Fleet is decreased below the current size of the EC2 Fleet. Valid values are `no-termination` and `termination`.

**LaunchTemplateId**

The ID of the launch template to use. You must specify either the launch template ID or launch template name. The launch template must specify an Amazon Machine Image (AMI). For more information about creating launch templates, see Launching an Instance from a Launch Template (p. 338).

**LaunchTemplateName**

The name of the launch template to use. You must specify either the launch template ID or launch template name. The launch template must specify an Amazon Machine Image (AMI). For more information, see Launching an Instance from a Launch Template (p. 338).

**Version**

The version number of the launch template.

**InstanceType**

(Optional) The instance type. If entered, this value overrides the launch template. The instance types must have the minimum hardware specifications that you need (vCPUs, memory, or storage).
MaxPrice
(Optional) The maximum price per unit hour that you are willing to pay for a Spot Instance. If entered, this value overrides the launch template. You can use the default maximum price (the On-Demand price) or specify the maximum price that you are willing to pay. Your Spot Instances are not launched if your maximum price is lower than the Spot price for the instance types that you specified.

SubnetId
(Optional) The ID of the subnet in which to launch the instances. If entered, this value overrides the launch template.

To create a new VPC, go the Amazon VPC console. When you are done, return to the JSON file and enter the new subnet ID.

AvailabilityZone
(Optional) The Availability Zone in which to launch the instances. The default is to let AWS choose the zones for your instances. If you prefer, you can specify specific zones. If entered, this value overrides the launch template.

Specify one or more Availability Zones. If you have more than one subnet in a zone, specify the appropriate subnet. To add subnets, go to the Amazon VPC console. When you are done, return to the JSON file and enter the new subnet ID.

WeightedCapacity
(Optional) The number of units provided by the specified instance type. If entered, this value overrides the launch template.

Priority
The priority for the launch template override. If AllocationStrategy is set to prioritized, EC2 Fleet uses priority to determine which launch template override to use first in fulfilling On-Demand capacity. The highest priority is launched first. Valid values are whole numbers starting at 0. The lower the number, the higher the priority. If no number is set, the override has the lowest priority.

TotalTargetCapacity
The number of instances to launch. You can choose instances or performance characteristics that are important to your application workload, such as vCPUs, memory, or storage. If the request type is maintain, you can specify a target capacity of 0 and add capacity later.

OnDemandTargetCapacity
(Optional) The number of On-Demand Instances to launch. This number must be less than the TotalTargetCapacity.

SpotTargetCapacity
(Optional) The number of Spot Instances to launch. This number must be less than the TotalTargetCapacity.

DefaultTargetCapacityType
If the value for TotalTargetCapacity is higher than the combined values for OnDemandTargetCapacity and SpotTargetCapacity, the difference is launched as the instance purchasing option specified here. Valid values are on-demand or spot.

TerminateInstancesWithExpiration
(Optional) By default, Amazon EC2 terminates your instances when the EC2 Fleet request expires. The default value is true. To keep them running after your request expires, do not enter a value for this parameter.
Type

(Optional) Indicates whether the EC2 Fleet submits a synchronous one-time request for your desired capacity (instant), or an asynchronous one-time request for your desired capacity, but with no attempt to maintain the capacity or to submit requests in alternative capacity pools if capacity is unavailable (request), or submits an asynchronous request for your desired capacity and continues to maintain your desired capacity by replenishing interrupted Spot Instances (maintain). Valid values are instant, request, and maintain. The default value is maintain. For more information, see EC2 Fleet Request Types (p. 354).

ValidFrom

(Optional) To create a request that is valid only during a specific time period, enter a start date.

ValidUntil

(Optional) To create a request that is valid only during a specific time period, enter an end date.

ReplaceUnhealthyInstances

(Optional) To replace unhealthy instances in an EC2 Fleet that is configured to maintain the fleet, enter true. Otherwise, leave this parameter empty.

TagSpecifications

(Optional) The key-value pair for tagging the EC2 Fleet request on creation. The value for ResourceType must be fleet, otherwise the fleet request fails. To tag instances at launch, specify the tags in the launch template (p. 340). For information about tagging after launch, see Tagging Your Resources (p. 966).

Creating an EC2 Fleet

When you create an EC2 Fleet, you must specify a launch template that includes information about the instances to launch, such as the instance type, Availability Zone, and the maximum price you are willing to pay.

You can create an EC2 Fleet that includes multiple launch specifications that override the launch template. The launch specifications can vary by instance type, Availability Zone, subnet, and maximum price, and can include a different weighted capacity.

When you create an EC2 Fleet, use a JSON file to specify information about the instances to launch. For more information, see EC2 Fleet JSON Configuration File Reference (p. 367).

EC2 Fleets can only be created using the AWS CLI.

To create an EC2 Fleet (AWS CLI)

- Use the following create-fleet (AWS CLI) command to create an EC2 Fleet.

```bash
aws ec2 create-fleet --cli-input-json file://file_name.json
```

For example configuration files, see EC2 Fleet Example Configurations (p. 376).

The following is example output for a fleet of type request or maintain.

```json
{
   "FleetId": "fleet-12a34b55-67cd-8ef9-ba9b-9208dEXAMPLE"
}
```

The following is example output for a fleet of type instant that launched the target capacity.
The following is example output for a fleet of type instant that launched part of the target capacity with errors for instances that were not launched.

```json
{
  "FleetId": "fleet-12a34b55-67cd-8ef9-ba9b-9208dEXAMPLE",
  "Errors": [
    {
      "LaunchTemplateAndOverrides": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-01234a567b8910abcEXAMPLE",
          "Version": "1"
        },
        "Overrides": {
          "InstanceType": "c4.xlarge",
          "AvailabilityZone": "us-east-1a"
        },
        "Lifecycle": "on-demand",
        "InstanceIds": [
          "i-5678901234abcdef0",
          "i-5432109876abcdef9"
        ],
        "InstanceType": "c4.large",
        "Platform": null
      },
      "LaunchTemplateAndOverrides": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-01234a567b8910abcEXAMPLE",
          "Version": "1"
        },
        "Overrides": {
          "InstanceType": "c4.large",
          "AvailabilityZone": "us-east-1a"
        },
        "Lifecycle": "on-demand",
        "InstanceIds": [
          "i-5678901234abcdef0",
          "i-5432109876abcdef9"
        ],
        "InstanceType": "c4.large",
        "Platform": null
      },
      "LaunchTemplateAndOverrides": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-01234a567b8910abcEXAMPLE",
          "Version": "1"
        },
        "Overrides": {
          "InstanceType": "c5.large",
          "AvailabilityZone": "us-east-1a"
        },
        "Lifecycle": "on-demand",
        "InstanceIds": [
          "i-1234567890abcdef0",
          "i-9876543210abcdef9"
        ],
        "InstanceType": "c5.large",
        "Platform": null
      }
    }
  ]
}
```
The following is example output for a fleet of type `instant` that launched no instances.

```json
{
  "FleetId": "fleet-12a34b55-67cd-8ef9-ba9b-9208dEXAMPLE",
  "Errors": [
    {
      "LaunchTemplateAndOverrides": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-01234a567b8910abcEXAMPLE",
          "Version": "1"
        },
        "Overrides": {
          "InstanceType": "c4.xlarge",
          "AvailabilityZone": "us-east-1a"
        },
        "Lifecycle": "on-demand",
        "InstanceIds": [
          "i-1234567890abcdef0",
          "i-9876543210abcdef9"
        ],
        "InstanceType": "c5.large",
        "Platform": null
      },
      "LaunchTemplateAndOverrides": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-01234a567b8910abcEXAMPLE",
          "Version": "1"
        },
        "Overrides": {
          "InstanceType": "c4.xlarge",
          "AvailabilityZone": "us-east-1a"
        },
        "Lifecycle": "on-demand",
        "ErrorCode": "InsufficientCapacity",
        "ErrorMessage": 
        "InstanceType": "c4.xlarge",
        "Platform": null
      },
      "LaunchTemplateAndOverrides": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-01234a567b8910abcEXAMPLE",
          "Version": "1"
        },
        "Overrides": {
          "InstanceType": "c5.large",
          "AvailabilityZone": "us-east-1a"
        },
        "Lifecycle": "on-demand",
        "ErrorCode": "InsufficientCapacity",
        "ErrorMessage": 
        "InstanceType": "c5.large",
        "Platform": null
      }
    }
  ]
}
```
Tagging an EC2 Fleet

To help categorize and manage your EC2 Fleet requests, you can tag them with custom metadata. For more information, see Tagging Your Amazon EC2 Resources (p. 964).

You can assign a tag to an EC2 Fleet request when you create it, or afterward. Tags assigned to the fleet request are not assigned to the instances launched by the fleet.

To tag a new EC2 Fleet request

To tag an EC2 Fleet request when you create it, specify the key-value pair in the JSON file (p. 366) used to create the fleet. The value for ResourceType must be fleet. If you specify another value, the fleet request fails.

To tag instances launched by an EC2 Fleet

To tag instances when they are launched by the fleet, specify the tags in the launch template (p. 340) that is referenced in the EC2 Fleet request.

To tag an existing EC2 Fleet request and instance (AWS CLI)

Use the following create-tags command to tag existing resources.

```bash
aws ec2 create-tags --resources fleet-12a34b55-67cd-8ef9-ba9b-9208dEXAMPLE i-1234567890abcdef0 --tags Key=purpose,Value=test
```

Monitoring Your EC2 Fleet

The EC2 Fleet launches On-Demand Instances when there is available capacity, and launches Spot Instances when your maximum price exceeds the Spot price and capacity is available. The On-Demand Instances run until you terminate them, and the Spot Instances run until they are interrupted or you terminate them.

The returned list of running instances is refreshed periodically and might be out of date.

To monitor your EC2 Fleet (AWS CLI)

Use the following describe-fleets command to describe your EC2 Fleets.

```bash
aws ec2 describe-fleets
```

The following is example output.

```json
{
  "Fleets": [
    {
      "Type": "maintain",
      "FulfilledCapacity": 2.0,
      "LaunchTemplateConfigs": [
        {
          "LaunchTemplateSpecification": {
            "Version": "2",
            "LaunchTemplateId": "lt-07b3bc7625cdab851"
          }
        }
      }
    }
  ]
}
```
Use the following `describe-fleet-instances` command to describe the instances for the specified EC2 Fleet.

```bash
aws ec2 describe-fleet-instances --fleet-id fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE
```

```
{
  "ActiveInstances": [
    {
      "InstanceId": "i-09cd595998cb3765e",
      "InstanceHealth": "healthy",
      "InstanceType": "m4.large",
      "SpotInstanceRequestId": "sir-86k84j6p"
    },
    {
      "InstanceId": "i-09cf95167ca219f17",
      "InstanceHealth": "healthy",
      "InstanceType": "m4.large",
      "SpotInstanceRequestId": "sir-dvxi7fsm"
    }
  ],
  "FleetId": "fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE"
}
```

Use the following `describe-fleet-history` command to describe the history for the specified EC2 Fleet for the specified time.

```bash
aws ec2 describe-fleet-history --fleet-request-id fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --start-time 2018-04-10T00:00:00Z
```

```
{
  "HistoryRecords": [],
  "FleetId": "fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE",
  "LastEvaluatedTime": "1970-01-01T00:00:00Z",
  "StartTime": "2018-04-09T23:53:20.000Z"
}
```
Modifying an EC2 Fleet

You can modify an EC2 Fleet that is in the submitted or active state. When you modify a fleet, it enters the modifying state.

You can modify the following parameters of an EC2 Fleet:

- **target-capacity-specification** – Increase or decrease the target capacity for TotalTargetCapacity, OnDemandTargetCapacity, and SpotTargetCapacity.
- **excess-capacity-termination-policy** – Whether running instances should be terminated if the total target capacity of the EC2 Fleet is decreased below the current size of the fleet. Valid values are no-termination and termination.

Note

You can only modify an EC2 Fleet that has Type=maintain.

When you increase the target capacity, the EC2 Fleet launches the additional instances according to the instance purchasing option specified for DefaultTargetCapacityType, which are either On-Demand Instances or Spot Instances.

If the DefaultTargetCapacityType is spot, the EC2 Fleet launches the additional Spot Instances according to its allocation strategy. If the allocation strategy is lowestPrice, the fleet launches the instances from the lowest-priced Spot Instance pool in the request. If the allocation strategy is diversified, the fleet distributes the instances across the pools in the request.

When you decrease the target capacity, the EC2 Fleet deletes any open requests that exceed the new target capacity. You can request that the fleet terminate instances until the size of the fleet reaches the new target capacity. If the allocation strategy is lowestPrice, the fleet terminates the instances with the highest price per unit. If the allocation strategy is diversified, the fleet terminates instances across the pools. Alternatively, you can request that EC2 Fleet keep the fleet at its current size, but not replace any Spot Instances that are interrupted or any instances that you terminate manually.

When an EC2 Fleet terminates a Spot Instance because the target capacity was decreased, the instance receives a Spot Instance interruption notice.

To modify an EC2 Fleet (AWS CLI)

Use the following modify-fleet command to update the target capacity of the specified EC2 Fleet.

```bash
aws ec2 modify-fleet --fleet-id fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --target-capacity-specification TotalTargetCapacity=20
```

If you are decreasing the target capacity but want to keep the fleet at its current size, you can modify the previous command as follows.

```bash
aws ec2 modify-fleet --fleet-id fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --target-capacity-specification TotalTargetCapacity=10 --excess-capacity-termination-policy no-termination
```

Deleting an EC2 Fleet

If you no longer require an EC2 Fleet, you can delete it. After you delete a fleet, it launches no new instances.

You must specify whether the EC2 Fleet must terminate its instances. If you specify that the instances must be terminated when the fleet is deleted, it enters the deleted_terminating state. Otherwise, it enters the deleted_running state, and the instances continue to run until they are interrupted or you terminate them manually.
To delete an EC2 Fleet (AWS CLI)

Use the `delete-fleets` command and the `--terminate-instances` parameter to delete the specified EC2 Fleet and terminate the instances.

```bash
aws ec2 delete-fleets --fleet-ids fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --terminate-instances
```

The following is example output.

```json
{
  "UnsuccessfulFleetDeletions": [],
  "SuccessfulFleetDeletions": [
    {
      "CurrentFleetState": "deleted_terminating",
      "PreviousFleetState": "active",
      "FleetId": "fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE"
    }
  ]
}
```

You can modify the previous command using the `--no-terminate-instances` parameter to delete the specified EC2 Fleet without terminating the instances.

```bash
aws ec2 delete-fleets --fleet-ids fleet-73fbd2ce-aa30-494c-8788-1cee4EXAMPLE --no-terminate-instances
```

The following is example output.

```json
{
  "UnsuccessfulFleetDeletions": [],
  "SuccessfulFleetDeletions": [
    {
      "CurrentFleetState": "deleted_running",
      "PreviousFleetState": "active",
      "FleetId": "fleet-4b8aaae8-dfb5-436d-a4c6-3dafa4c6b7dcEXAMPLE"
    }
  ]
}
```

**EC2 Fleet Example Configurations**

The following examples show launch configurations that you can use with the `create-fleet` command to create an EC2 Fleet. For more information, see the EC2 Fleet JSON Configuration File Reference (p. 367).

1. Launch Spot Instances as the default purchasing option (p. 376)
2. Launch On-Demand Instances as the default purchasing option (p. 377)
3. Launch On-Demand Instances as the primary capacity (p. 377)
4. Launch Spot Instances using the lowestPrice allocation strategy (p. 378)

**Example 1: Launch Spot Instances as the Default Purchasing Option**

The following example specifies the minimum parameters required in an EC2 Fleet: a launch template, target capacity, and default purchasing option. The launch template is identified by its launch template ID and version number. The target capacity for the fleet is 2 instances, and the default purchasing option is `spot`, which results in the fleet launching 2 Spot Instances.
Example 2: Launch On-Demand Instances as the Default Purchasing Option

The following example specifies the minimum parameters required in an EC2 Fleet: a launch template, target capacity, and default purchasing option. The launch template is identified by its launch template ID and version number. The target capacity for the fleet is 2 instances, and the default purchasing option is on-demand, which results in the fleet launching 2 On-Demand Instances.

```json
{
    "LaunchTemplateConfigs": [
        {
            "LaunchTemplateSpecification": {
                "LaunchTemplateId": "lt-0e8c754449b27161c",
                "Version": "1"
            }
        },
        {
            "TargetCapacitySpecification": {
                "TotalTargetCapacity": 2,
                "DefaultTargetCapacityType": "on-demand"
            }
        }
    ]
}
```

Example 3: Launch On-Demand Instances as the Primary Capacity

The following example specifies the total target capacity of 2 instances for the fleet, and a target capacity of 1 On-Demand Instance. The default purchasing option is spot. The fleet launches 1 On-Demand instance as specified, but needs to launch one more instance to fulfill the total target capacity. The purchasing option for the difference is calculated as TotalTargetCapacity - OnDemandTargetCapacity = DefaultTargetCapacityType, which results in the fleet launching 1 Spot Instance.

```json
{
    "LaunchTemplateConfigs": [
        {
            "LaunchTemplateSpecification": {
                "LaunchTemplateId": "lt-0e8c754449b27161c",
                "Version": "1"
            }
        },
        {
            "TargetCapacitySpecification": {
                "TotalTargetCapacity": 2,
                "OnDemandTargetCapacity": 1,
                "DefaultTargetCapacityType": "spot"
            }
        }
    ]
}
```
Example 4: Launch Spot Instances Using the Lowest Price Allocation Strategy

If the allocation strategy for Spot Instances is not specified, the default allocation strategy, which is `lowestPrice`, is used. The following example uses the `lowestPrice` allocation strategy. The three launch specifications, which override the launch template, have different instance types but the same weighted capacity and subnet. The total target capacity is 2 instances and the default purchasing option is `spot`. The EC2 Fleet launches 2 Spot Instances using the instance type of the launch specification with the lowest price.

```json
{
    "LaunchTemplateConfigs": [
        {
            "LaunchTemplateSpecification": {
                "LaunchTemplateId": "lt-0e8c754449b27161c",
                "Version": "1"
            }
        }
    ],
    "Overrides": [
        {
            "InstanceType": "c4.large",
            "WeightedCapacity": 1,
            "SubnetId": "subnet-a4f6c5d3"
        },
        {
            "InstanceType": "c3.large",
            "WeightedCapacity": 1,
            "SubnetId": "subnet-a4f6c5d3"
        },
        {
            "InstanceType": "c5.large",
            "WeightedCapacity": 1,
            "SubnetId": "subnet-a4f6c5d3"
        }
    ],
    "TargetCapacitySpecification": {
        "TotalTargetCapacity": 2,
        "DefaultTargetCapacityType": "spot"
    }
}
}
Connecting to Your Windows Instance

Amazon EC2 instances created from most Windows Amazon Machine Images (AMIs) enable you to connect using Remote Desktop. Remote Desktop uses the Remote Desktop Protocol (RDP) and enables you to connect to and use your instance in the same way you use a computer sitting in front of you. It is available on most editions of Windows and available for Mac OS.

For information about connecting to a Linux instance, see Connect to Your Linux Instance in the Amazon EC2 User Guide for Linux Instances.

Contents
- Prerequisites (p. 379)
- Connect to Your Windows Instance (p. 380)
- Connect to a Windows Instance Using Its IPv6 Address (p. 381)
- Connecting to a Windows Instance Using AWS Systems Manager Session Manager (p. 382)
- Transfer Files to Windows Instances (p. 382)

Prerequisites

- Install an RDP client
  - [Windows] Windows includes an RDP client by default. To verify, type mstsc at a Command Prompt window. If your computer doesn't recognize this command, see the Windows home page and search for the download for the Microsoft Remote Desktop app.
  - [Mac OS X] Download the Microsoft Remote Desktop app from the Mac App Store.
  - [Linux] Use rdesktop.
- Get the ID of the instance.
  You can get the ID of your instance using the Amazon EC2 console (from the Instance ID column). If you prefer, you can use the describe-instances (AWS CLI) or Get-EC2Instance (AWS Tools for Windows PowerShell) command.
- Get the public DNS name of the instance.
  You can get the public DNS for your instance using the Amazon EC2 console. Check the Public DNS (IPv4) column. If this column is hidden, choose the Show/Hide icon and select Public DNS (IPv4). If you prefer, you can use the describe-instances (AWS CLI) or Get-EC2Instance (AWS Tools for Windows PowerShell) command.
- (IPv6 only) Get the IPv6 address of the instance.
  If you've assigned an IPv6 address to your instance, you can optionally connect to the instance using its IPv6 address instead of a public IPv4 address or public IPv4 DNS hostname. Your local computer must have an IPv6 address and must be configured to use IPv6. You can get the IPv6 address of your instance using the Amazon EC2 console. Check the IPv6 IPs field. If you prefer, you can use the describe-instances (AWS CLI) or Get-EC2Instance (AWS Tools for Windows PowerShell) command. For more information about IPv6, see IPv6 Addresses (p. 658).
- Locate the private key
  Get the fully-qualified path to the location on your computer of the .pem file for the key pair that you specified when you launched the instance. For more information about how you created your key pair, see Creating a Key Pair Using Amazon EC2.
- Enable inbound RDP traffic from your IP address to your instance
Ensure that the security group associated with your instance allows incoming RDP traffic from your IP address. The default security group does not allow incoming RDP traffic by default. For more information, see Authorizing Inbound Traffic for Your Windows Instances (p. 654).

- For the best experience using Internet Explorer, run the latest version.

**Connect to Your Windows Instance**

To connect to a Windows instance, you must retrieve the initial administrator password (see step 2 below) and then specify this password when you connect to your instance using Remote Desktop.

The name of the administrator account depends on the language of the operating system. For example, for English, it's Administrator, for French it's Administrateur, and for Portuguese it's Administrador. For more information, see Localized Names for Administrator Account in Windows in the Microsoft TechNet Wiki.

If you've joined your instance to a domain, you can connect to your instance using domain credentials you've defined in AWS Directory Service. On the Remote Desktop login screen, instead of using the local computer name and the generated password, use the fully-qualified user name for the administrator (for example, corp.example.com\Admin) and the password for this account.

The license for the Windows Server operating system (OS) allows two simultaneous remote connections for administrative purposes. The license for Windows Server is included in the price of your Windows instance. If you need more than two simultaneous remote connections, you must purchase a Remote Desktop Services (RDS) license. If you attempt a third connection, an error occurs. For more information, see Configure the Number of Simultaneous Remote Connections Allowed for a Connection.

**To connect to your Windows instance using an RDP client**

1. In the Amazon EC2 console, select the instance, and then choose Connect.
2. In the Connect To Your Instance dialog box, choose Get Password (it will take a few minutes after the instance is launched before the password is available).
3. Choose Browse and navigate to the private key file you created when you launched the instance. Select the file and choose Open to copy the entire contents of the file into the Contents field.
4. Choose Decrypt Password. The console displays the default administrator password for the instance in the Connect To Your Instance dialog box, replacing the link to Get Password shown previously with the actual password.
5. Record the default administrator password, or copy it to the clipboard. You need this password to connect to the instance.
6. Choose Download Remote Desktop File. Your browser prompts you to either open or save the .rdp file. Either option is fine. When you have finished, you can choose Close to dismiss the Connect To Your Instance dialog box.

   - If you opened the .rdp file, you'll see the Remote Desktop Connection dialog box.
   - If you saved the .rdp file, navigate to your downloads directory, and open the .rdp file to display the dialog box.
7. You may get a warning that the publisher of the remote connection is unknown. You can continue to connect to your instance.
8. When prompted, log in to the instance, using the administrator account for the operating system and the password that you recorded or copied previously. If your Remote Desktop Connection already has an administrator account set up, you might have to choose the Use another account option and type the user name and password manually.

**Note**

Sometimes copying and pasting content can corrupt data. If you encounter a “Password Failed” error when you log in, try typing in the password manually.
9. Due to the nature of self-signed certificates, you may get a warning that the security certificate could not be authenticated. Use the following steps to verify the identity of the remote computer, or simply choose Yes or Continue to continue if you trust the certificate.

   a. If you are using Remote Desktop Connection from a Windows PC, choose View certificate. If you are using Microsoft Remote Desktop on a Mac, choose Show Certificate.

   b. Choose the Details tab, and scroll down to the Thumbprint entry on a Windows PC, or the SHA1 Fingerprints entry on a Mac. This is the unique identifier for the remote computer’s security certificate.

   c. In the Amazon EC2 console, select the instance, choose Actions, and then choose Get System Log.

   d. In the system log output, look for an entry labeled RDPCERTIFICATE-THUMBPRINT. If this value matches the thumbprint or fingerprint of the certificate, you have verified the identity of the remote computer.

   e. If you are using Remote Desktop Connection from a Windows PC, return to the Certificate dialog box and choose OK. If you are using Microsoft Remote Desktop on a Mac, return to the Verify Certificate and choose Continue.

   f. [Windows] Choose Yes in the Remote Desktop Connection window to connect to your instance.

   [Mac OS] Log in as prompted, using the default administrator account and the default administrator password that you recorded or copied previously. Note that you might need to switch spaces to see the login screen. For more information about spaces, see support.apple.com/en-us/HT204100.

   g. If you receive an error while attempting to connect to your instance, see Remote Desktop Can't Connect to the Remote Computer (p. 1045).

After you connect, we recommend that you do the following:

- Change the administrator password from the default value. You change the password while logged on to the instance itself, just as you would on any other Windows Server.

- Create another user account with administrator privileges on the instance. Another account with administrator privileges is a safeguard if you forget the administrator password or have a problem with the administrator account. The user account must have permission to access the instance remotely. Open System Properties by right-clicking on the This PC icon on your Windows desktop or File Explorer and selecting Properties. Choose Remote settings, and choose Select Users to add the user to the Remote Desktop Users group.

### Connect to a Windows Instance Using Its IPv6 Address

If you've enabled your VPC for IPv6 and assigned an IPv6 address to your Windows instance, you can use an RDP client to connect to your instance using its IPv6 address instead of a public IPv4 address or public DNS hostname. For more information, see IPv6 Addresses (p. 658).

**To connect to your Windows instance using its IPv6 address**

1. In the Amazon EC2 console, select the instance, and then choose Connect.

2. In the Connect To Your Instance dialog box, choose Get Password (it will take a few minutes after the instance is launched before the password is available).

3. Choose Browse and navigate to the private key file you created when you launched the instance. Select the file and choose Open to copy the entire contents of the file into the Contents field.

4. Choose Decrypt Password.

5. Copy the default administrator password. You need this password to connect to the instance.

6. Open the RDP client on your computer.
7. [Windows] For the RDP client on a Windows computer, choose **Show Options** and do the following:
   - For **Computer**, type the IPv6 address of your Windows instance, for example, 2001:db8:1234:1a00:9691:9503:25ad:1761.
   - For **User name**, enter **Administrator**.
   - Choose **Connect**.

    [Mac OS X] For the Microsoft Remote Desktop app, choose **New** and do the following:
    - For **PC Name**, enter the IPv6 address of your Windows instance; for example, 2001:db8:1234:1a00:9691:9503:25ad:1761.
    - For **User name**, enter **Administrator**.
    - Close the dialog box. Under **My Desktops**, select the connection and choose **Start**.

8. Due to the nature of self-signed certificates, you may get a warning that the security certificate could not be authenticated. Use the following steps to verify the identity of the remote computer, or simply choose **Yes** or **Continue** to continue if you trust the certificate.

9. When prompted, enter the password that you recorded or copied previously.

---

### Connecting to a Windows Instance Using AWS Systems Manager Session Manager

AWS Systems Manager Session Manager is a fully managed AWS Systems Manager capability that lets you manage your Amazon EC2 instances through an interactive one-click browser-based shell or through the AWS CLI. For more information about how to connect to your instance with SSM Session Manager, see [AWS Systems Manager Session Manager](https://docs.aws.amazon.com/systems-manager/latest/userguide/ssm-cli-ref.html) in the *Systems Manager User Guide*.

### Transfer Files to Windows Instances

You can work with your Windows instance the same way that you would work with any Windows server. For example, you can transfer files between a Windows instance and your local computer using the local file sharing feature of the Microsoft Remote Desktop Connection software. If you enable this option, you can access your local files from your Windows instances. You can access local files on hard disk drives, DVD drives, portable media drives, and mapped network drives.

To make local devices and resources available to a remote session on Windows, map the remote session drive to your local drive:

1. Open the Remote Desktop Connection client.
2. Choose **Show Options**.
3. Select the **Local Resources** tab.
4. Under the Local Devices and resources section, choose **More**.
5. Open Drives and select the local drive to map to your Windows instance.
6. Choose **OK**.
7. Choose **Connect** to connect to your Windows instance.

For more information on making local devices available to a remote session on a Mac computer, see:

[Get Started with Remote Desktop on Mac](https://docs.aws.amazon.com/systems-manager/latest/userguide/remote-desktop.html)
Stop and Start Your Instance

You can stop and restart your instance if it has an Amazon EBS volume as its root device. The instance retains its instance ID, but can change as described in the Overview (p. 383) section.

When you stop an instance, we shut it down. We don't charge usage for a stopped instance, or data transfer fees, but we do charge for the storage for any Amazon EBS volumes. Each time you start a stopped instance we charge a full instance hour, even if you make this transition multiple times within a single hour.

While the instance is stopped, you can treat its root volume like any other volume, and modify it (for example, repair file system problems or update software). You just detach the volume from the stopped instance, attach it to a running instance, make your changes, detach it from the running instance, and then reattach it to the stopped instance. Make sure that you reattach it using the storage device name that's specified as the root device in the block device mapping for the instance.

If you decide that you no longer need an instance, you can terminate it. As soon as the state of an instance changes to shutting-down or terminated, we stop charging for that instance. For more information, see Terminate Your Instance (p. 388). If you'd rather hibernate the instance, see Hibernate Your Instance (p. 385). For more information, see Differences Between Reboot, Stop, Hibernate, and Terminate (p. 332).

Contents

- Overview (p. 383)
- What Happens When You Stop an Instance (API) (p. 384)
- Stopping and Starting Your Instances (p. 384)
- Modifying a Stopped Instance (p. 385)
- Troubleshooting (p. 385)

Overview

When you stop a running instance, the following happens:

- The instance performs a normal shutdown and stops running; its status changes to stopping and then stopped.
- Any Amazon EBS volumes remain attached to the instance, and their data persists.
- Any data stored in the RAM of the host computer or the instance store volumes of the host computer is gone.
- In most cases, the instance is migrated to a new underlying host computer when it's started.
- The instance retains its private IPv4 addresses and any IPv6 addresses when stopped and restarted. We release the public IPv4 address and assign a new one when you restart it.
- The instance retains its associated Elastic IP addresses. You're charged for any Elastic IP addresses associated with a stopped instance. With EC2-Classic, an Elastic IP address is dissociated from your instance when you stop it. For more information, see EC2-Classic (p. 752).
- When you stop and start a Windows instance, the EC2Config service performs tasks on the instance, such as changing the drive letters for any attached Amazon EBS volumes. For more information about these defaults and how you can change them, see Configuring a Windows Instance Using the EC2Config Service (p. 405) in the Amazon EC2 User Guide for Windows Instances.
- If your instance is in an Auto Scaling group, the Amazon EC2 Auto Scaling service marks the stopped instance as unhealthy, and may terminate it and launch a replacement instance. For more information, see Health Checks for Auto Scaling Instances in the Amazon EC2 Auto Scaling User Guide.
• When you stop a ClassicLink instance, it’s unlinked from the VPC to which it was linked. You must link the instance to the VPC again after restarting it. For more information about ClassicLink, see ClassicLink (p. 760).

For more information, see Differences Between Reboot, Stop, Hibernate, and Terminate (p. 332).

You can modify the following attributes of an instance only when it is stopped:

• Instance type
• User data
• Kernel
• RAM disk

If you try to modify these attributes while the instance is running, Amazon EC2 returns the IncorrectInstanceState error.

**What Happens When You Stop an Instance (API)**

When an EC2 instance is stopped using the `stop-instances` command, the following is registered at the OS level:

• The API request will send a button press event to the guest.
• Various system services will be stopped as a result of the button press event. `systemd` handles a graceful shutdown of the system. Graceful shutdown is triggered by the ACPI shutdown button press event from the hypervisor.
• ACPI shutdown will be initiated.
• The instance will shut down when the graceful shutdown process exits. There is no configurable OS shutdown time.

**Stopping and Starting Your Instances**

You can start and stop your Amazon EBS-backed instance using the console or the command line.

By default, when you initiate a shutdown from an Amazon EBS-backed instance (using the `shutdown` or `poweroff` command), the instance stops. You can change this behavior so that it terminates instead. For more information, see Changing the Instance Initiated Shutdown Behavior (p. 391).

**To stop and start an Amazon EBS-backed instance using the console**

1. In the navigation pane, choose **Instances**, and select the instance.
2. Choose **Actions**, select **Instance State**, and then choose **Stop**. If **Stop** is disabled, either the instance is already stopped or its root device is an instance store volume.
   **Warning**
   When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.
3. In the confirmation dialog box, choose **Yes, Stop**. It can take a few minutes for the instance to stop.
4. While your instance is stopped, you can modify certain instance attributes. For more information, see Modifying a Stopped Instance (p. 385).
5. To restart the stopped instance, select the instance, and choose **Actions, Instance State, Start**.
6. In the confirmation dialog box, choose **Yes, Start**. It can take a few minutes for the instance to enter the running state.
To stop and start an Amazon EBS-backed instance using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- stop-instances and start-instances (AWS CLI)
- Stop-EC2Instance and Start-EC2Instance (AWS Tools for Windows PowerShell)

Modifying a Stopped Instance

You can change the instance type, user data, and EBS-optimization attributes of a stopped instance using the AWS Management Console or the command line interface. You can't use the AWS Management Console to modify the DeleteOnTermination, kernel, or RAM disk attributes.

To modify an instance attribute

- To change the instance type, see Changing the Instance Type (p. 188).
- To change the user data for your instance, see Working with Instance User Data (p. 480).
- To enable or disable EBS-optimization for your instance, see Modifying EBS-Optimization (p. 886).
- To change the DeleteOnTermination attribute of the root volume for your instance, see Updating the Block Device Mapping of a Running Instance (p. 937).

To modify an instance attribute using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- modify-instance-attribute (AWS CLI)
- Edit-EC2InstanceAttribute (AWS Tools for Windows PowerShell)

Troubleshooting

If you have stopped your Amazon EBS-backed instance and it appears "stuck" in the stopping state, you can forcibly stop it. For more information, see Troubleshooting Stopping Your Instance in the Amazon EC2 User Guide for Linux Instances.

Hibernate Your Instance

When you hibernate an instance, we signal the operating system to perform hibernation (suspend-to-disk), which saves the contents from the instance memory (RAM) to your Amazon EBS root volume. We persist the instance's Amazon EBS root volume and any attached Amazon EBS data volumes. When you restart your instance, the Amazon EBS root volume is restored to its previous state, the RAM contents are reloaded, and the processes that were previously running on the instance are resumed. Previously attached data volumes are reattached and the instance retains its instance ID. For more information, see Hibernate Your Instance in the Amazon EC2 User Guide for Linux Instances.

Important

Hibernation is currently not supported on Windows instances.

Reboot Your Instance

An instance reboot is equivalent to an operating system reboot. In most cases, it takes only a few minutes to reboot your instance. When you reboot an instance, it remains on the same physical host, so
your instance keeps its public DNS name (IPv4), private IPv4 address, IPv6 address (if applicable), and any
data on its instance store volumes.

Rebooting an instance doesn't start a new instance billing hour, unlike stopping and restarting your
instance.

We might schedule your instance for a reboot for necessary maintenance, such as to apply updates that
require a reboot. No action is required on your part; we recommend that you wait for the reboot to occur
within its scheduled window. For more information, see Scheduled Events for Your Instances (p. 539).

We recommend that you use the Amazon EC2 console, a command line tool, or the Amazon EC2 API to
reboot your instance instead of running the operating system reboot command from your instance. If
you use the Amazon EC2 console, a command line tool, or the Amazon EC2 API to reboot your instance,
we perform a hard reboot if the instance does not cleanly shut down within four minutes. If you use
AWS CloudTrail, then using Amazon EC2 to reboot your instance also creates an API record of when your
instance was rebooted.

If Windows is installing updates on your instance, we recommend that you do not reboot or shut down
your instance using the Amazon EC2 console or the command line until all the updates are installed.
When you use the Amazon EC2 console or the command line to reboot or shut down your instance, there
is a risk that your instance will be hard rebooted. A hard reboot while updates are being installed could
throw your instance into an unstable state.

**To reboot an instance using the console**

1. Open the Amazon EC2 console.
2. In the navigation pane, choose **Instances**.
3. Select the instance and choose **Actions, Instance State, Reboot**.
4. Choose **Yes, Reboot** when prompted for confirmation.

**To reboot an instance using the command line**

You can use one of the following commands. For more information about these command line interfaces,
see Accessing Amazon EC2 (p. 3).

- **reboot-instances** (AWS CLI)
- **Restart-EC2Instance** (AWS Tools for Windows PowerShell)

**Instance Retirement**

An instance is scheduled to be retired when AWS detects irreparable failure of the underlying hardware
hosting the instance. When an instance reaches its scheduled retirement date, it is stopped or terminated
by AWS. If your instance root device is an Amazon EBS volume, the instance is stopped, and you can start
it again at any time. Starting the stopped instance migrates it to new hardware. If your instance root
device is an instance store volume, the instance is terminated, and cannot be used again.

**Contents**

- Identifying Instances Scheduled for Retirement (p. 387)
- Working with Instances Scheduled for Retirement (p. 387)

For more information about types of instance events, see Scheduled Events for Your Instances (p. 539).
Identifying Instances Scheduled for Retirement

If your instance is scheduled for retirement, you'll receive an email prior to the event with the instance ID and retirement date. This email is sent to the address that's associated with your account; the same email address that you use to log in to the AWS Management Console. If you use an email account that you do not check regularly, then you can use the Amazon EC2 console or the command line to determine if any of your instances are scheduled for retirement. To update the contact information for your account, go to the Account Settings page.

To identify instances scheduled for retirement using the console

1. Open the Amazon EC2 console.
2. In the navigation pane, choose EC2 Dashboard. Under Scheduled Events, you can see the events associated with your Amazon EC2 instances and volumes, organized by region.

   Scheduled Events
   
   US East (N. Virginia):
   1 instances have scheduled events

3. If you have an instance with a scheduled event listed, select its link below the region name to go to the Events page.
4. The Events page lists all resources with events associated with them. To view instances that are scheduled for retirement, select Instance resources from the first filter list, and then Instance stop or retirement from the second filter list.
5. If the filter results show that an instance is scheduled for retirement, select it, and note the date and time in the Start time field in the details pane. This is your instance retirement date.

To identify instances scheduled for retirement using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- describe-instance-status (AWS CLI)
- Get-EC2InstanceStatus (AWS Tools for Windows PowerShell)

Working with Instances Scheduled for Retirement

There are a number of actions available to you when your instance is scheduled for retirement. The action you take depends on whether your instance root device is an Amazon EBS volume, or an instance store volume. If you do not know what your instance root device type is, you can find out using the Amazon EC2 console or the command line.

Determining Your Instance Root Device Type

To determine your instance root device type using the console

1. In the navigation pane, select Events. Use the filter lists to identify retiring instances, as demonstrated in the procedure above, Identifying instances scheduled for retirement (p. 387).
2. In the Resource Id column, select the instance ID to go to the Instances page.
3. Select the instance and locate the **Root device type** field in the **Description** tab. If the value is **ebs**, then your instance is EBS-backed. If the value is **instance-store**, then your instance is instance store-backed.

### To determine your instance root device type using the command line

You can use one of the following commands. For more information about these command line interfaces, see [Accessing Amazon EC2](p. 3).

- `describe-instances` (AWS CLI)
- `Get-EC2Instance` (AWS Tools for Windows PowerShell)

### Managing Instances Scheduled for Retirement

You can perform one of the actions listed below in order to preserve the data on your retiring instance. It's important that you take this action before the instance retirement date to prevent unforeseen downtime and data loss. For your EBS-backed instance, you can perform one of the following actions.

- Create an EBS-backed AMI from your instance so that you have a backup. Wait for the scheduled retirement date - when the instance is stopped - or stop the instance yourself before the retirement date. You can start the instance again at any time. For more information about stopping and starting your instance, and what to expect when your instance is stopped, such as the effect on public, private, and Elastic IP addresses associated with your instance, see [Stop and Start Your Instance](p. 383).
- Create an EBS-backed AMI from your instance, and launch a replacement instance. For more information, see [Creating a Custom Windows AMI](p. 64).

### Terminate Your Instance

You can delete your instance when you no longer need it. This is referred to as **terminating** your instance. As soon as the state of an instance changes to **shutting-down** or **terminated**, you stop incurring charges for that instance.

You can't connect to or restart an instance after you've terminated it. However, you can launch additional instances using the same AMI. If you'd rather stop and restart your instance, or hibernate it, see [Stop and Start Your Instance](p. 383) or [Hibernate Your Instance](p. 385). For more information, see [Differences Between Reboot, Stop, Hibernate, and Terminate](p. 332).

### Contents

- Instance Termination (p. 388)
- What Happens When You Terminate an Instance (API) (p. 389)
- Terminating an Instance (p. 389)
- Enabling Termination Protection for an Instance (p. 390)
- Changing the Instance Initiated Shutdown Behavior (p. 391)
- Preserving Amazon EBS Volumes on Instance Termination (p. 392)

### Instance Termination

After you terminate an instance, it remains visible in the console for a short while, and then the entry is automatically deleted. You cannot delete the terminated instance entry yourself. After an instance is terminated, resources such as tags and volumes are gradually disassociated from the instance, therefore may no longer be visible on the terminated instance after a short while.
When an instance terminates, the data on any instance store volumes associated with that instance is deleted.

By default, Amazon EBS root device volumes are automatically deleted when the instance terminates. However, by default, any additional EBS volumes that you attach at launch, or any EBS volumes that you attach to an existing instance persist even after the instance terminates. This behavior is controlled by the volume's `DeleteOnTermination` attribute, which you can modify. For more information, see Preserving Amazon EBS Volumes on Instance Termination (p. 392).

You can prevent an instance from being terminated accidentally by someone using the AWS Management Console, the CLI, and the API. This feature is available for both Amazon EC2 instance store-backed and Amazon EBS-backed instances. Each instance has a `DisableApiTermination` attribute with the default value of `false` (the instance can be terminated through Amazon EC2). You can modify this instance attribute while the instance is running or stopped (in the case of Amazon EBS-backed instances). For more information, see Enabling Termination Protection for an Instance (p. 390).

You can control whether an instance should stop or terminate when shutdown is initiated from the instance using an operating system command for system shutdown. For more information, see Changing the Instance InitiatedShutdown Behavior (p. 391).

If you run a script on instance termination, your instance might have an abnormal termination, because we have no way to ensure that shutdown scripts run. Amazon EC2 attempts to shut an instance down cleanly and run any system shutdown scripts; however, certain events (such as hardware failure) may prevent these system shutdown scripts from running.

What Happens When You Terminate an Instance (API)

When an EC2 instance is terminated using the `terminate-instances` command, the following is registered at the OS level:

- The API request will send a button press event to the guest.
- Various system services will be stopped as a result of the button press event. `systemd` handles a graceful shutdown of the system. Graceful shutdown is triggered by the ACPI shutdown button press event from the hypervisor.
- ACPI shutdown will be initiated.
- The instance will shut down when the graceful shutdown process exits. There is no configurable OS shutdown time.

Terminating an Instance

You can terminate an instance using the AWS Management Console or the command line.

To terminate an instance using the console

1. Before you terminate the instance, verify that you won't lose any data by checking that your Amazon EBS volumes won't be deleted on termination and that you've copied any data that you need from your instance store volumes to Amazon EBS or Amazon S3.
2. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
3. In the navigation pane, choose **Instances**.
4. Select the instance, and choose **Actions, Instance State, Terminate**.
5. Choose Yes, **Terminate** when prompted for confirmation.

To terminate an instance using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
• terminate-instances (AWS CLI)
• Stop-EC2Instance (AWS Tools for Windows PowerShell)

Enabling Termination Protection for an Instance

By default, you can terminate your instance using the Amazon EC2 console, command line interface, or API. If you want to prevent your instance from being accidentally terminated using Amazon EC2, you can enable termination protection for the instance. The DisableApiTermination attribute controls whether the instance can be terminated using the console, CLI, or API. By default, termination protection is disabled for your instance. You can set the value of this attribute when you launch the instance, while the instance is running, or while the instance is stopped (for Amazon EBS-backed instances).

The DisableApiTermination attribute does not prevent you from terminating an instance by initiating shutdown from the instance (using an operating system command for system shutdown) when the InstanceInitiatedShutdownBehavior attribute is set. For more information, see Changing the Instance Initiated Shutdown Behavior (p. 391).

Limits

You can't enable termination protection for Spot Instances—a Spot Instance is terminated when the Spot price exceeds the amount you're willing to pay for Spot Instances. However, you can prepare your application to handle Spot Instance interruptions. For more information, see Spot Instance Interruptions (p. 285).

The DisableApiTermination attribute does not prevent Amazon EC2 Auto Scaling from terminating an instance. For instances in an Auto Scaling group, use the following Amazon EC2 Auto Scaling features instead of Amazon EC2 termination protection:

• To prevent instances that are part of an Auto Scaling group from terminating on scale in, use instance protection. For more information, see Instance Protection in the Amazon EC2 Auto Scaling User Guide.
• To prevent Amazon EC2 Auto Scaling from terminating unhealthy instances, suspend the ReplaceUnhealthy process. For more information, see Suspending and Resuming Scaling Processes in the Amazon EC2 Auto Scaling User Guide.
• To specify which instances Amazon EC2 Auto Scaling should terminate first, choose a termination policy. For more information, see Customizing the Termination Policy in the Amazon EC2 Auto Scaling User Guide.

To enable termination protection for an instance at launch time

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the dashboard, choose Launch Instance and follow the directions in the wizard.
3. On the Configure Instance Details page, select the Enable termination protection check box.

To enable termination protection for a running or stopped instance

1. Select the instance, choose Actions, Instance Settings, and then choose Change Termination Protection.
2. Select Yes, Enable.

To disable termination protection for a running or stopped instance

1. Select the instance, choose Actions, Instance Settings, and then choose Change Termination Protection.
2. Select Yes, Disable.
To enable or disable termination protection using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- modify-instance-attribute (AWS CLI)
- Edit-EC2InstanceAttribute (AWS Tools for Windows PowerShell)

Changing the Instance Initiated Shutdown Behavior

By default, when you initiate a shutdown from an Amazon EBS-backed instance (using a command such as `shutdown` or `poweroff`), the instance stops (Note that `halt` does not issue a `poweroff` command and, if used, the instance will not terminate; instead, it will place the CPU into HLT and the instance will remain running). You can change this behavior using the `InstanceInitiatedShutdownBehavior` attribute for the instance so that it terminates instead. You can update this attribute while the instance is running or stopped.

You can update the `InstanceInitiatedShutdownBehavior` attribute using the Amazon EC2 console or the command line. The `InstanceInitiatedShutdownBehavior` attribute only applies when you perform a shutdown from the operating system of the instance itself; it does not apply when you stop an instance using the `StopInstances` API or the Amazon EC2 console.

To change the shutdown behavior of an instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance, select Actions, Instance Settings, and then choose Change Shutdown Behavior. The current behavior is already selected.
4. To change the behavior, select an option from the Shutdown behavior list, and then select Apply.

To change the shutdown behavior of an instance using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- modify-instance-attribute (AWS CLI)
- Edit-EC2InstanceAttribute (AWS Tools for Windows PowerShell)
Preserving Amazon EBS Volumes on Instance Termination

When an instance terminates, Amazon EC2 uses the value of the \texttt{DeleteOnTermination} attribute for each attached Amazon EBS volume to determine whether to preserve or delete the volume. The default value for the \texttt{DeleteOnTermination} attribute differs depending on whether or not the volume is a root volume of an instance.

By default, the \texttt{DeletionOnTermination} attribute for the root volume of an instance is set to \texttt{true}. Therefore, the default is to delete the root volume of an instance when the instance terminates. The \texttt{DeletionOnTermination} attribute can be set by the creator of an AMI as well as by the person who launches an instance. When the attribute is changed by the creator of an AMI or by the person who launches an instance, the new setting overrides the original AMI default setting. We recommend that you verify the default setting for the \texttt{DeletionOnTermination} attribute after you launch an instance with an AMI.

By default, when you attach an EBS volume to an instance, its \texttt{DeleteOnTermination} attribute is set to \texttt{false}. Therefore, the default is to preserve these volumes. You must delete a volume to avoid incurring further charges. For more information, see Deleting an Amazon EBS Volume (p. 818). After the instance terminates, you can take a snapshot of the preserved volume or attach it to another instance.

To verify the value of the \texttt{DeletionOnTermination} attribute for an EBS volume that is in-use, look at the instance's block device mapping. For more information, see Viewing the EBS Volumes in an Instance Block Device Mapping (p. 938).

You can change value of the \texttt{DeleteOnTermination} attribute for a volume when you launch the instance or while the instance is running.

**Examples**
- Changing the Root Volume to Persist at Launch Using the Console (p. 392)
- Changing the Root Volume to Persist at Launch Using the Command Line (p. 392)
- Changing the Root Volume of a Running Instance to Persist Using the Command Line (p. 393)

Changing the Root Volume to Persist at Launch Using the Console

Using the console, you can change the \texttt{DeleteOnTermination} attribute when you launch an instance. To change this attribute for a running instance, you must use the command line.

**To change the root volume of an instance to persist at launch using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the console dashboard, select \texttt{Launch Instance}.
3. On the \texttt{Choose an Amazon Machine Image (AMI)} page, choose an AMI and choose \texttt{Select}.
4. Follow the wizard to complete the \texttt{Choose an Instance Type} and \texttt{Configure Instance Details} pages.
5. On the \texttt{Add Storage} page, deselect the \texttt{Delete On Termination} check box for the root volume.
6. Complete the remaining wizard pages, and then choose \texttt{Launch}.

You can verify the setting by viewing details for the root device volume on the instance's details pane. Next to \texttt{Block devices}, click the entry for the root device volume. By default, \texttt{Delete on termination} is \texttt{True}. If you change the default behavior, \texttt{Delete on termination} is \texttt{False}.

Changing the Root Volume to Persist at Launch Using the Command Line

When you launch an EBS-backed instance, you can use one of the following commands to change the root device volume to persist. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
• **run-instances** (AWS CLI)
• **New-EC2Instance** (AWS Tools for Windows PowerShell)

For example, add the following option to your `run-instances` command:

```
--block-device-mappings file://mapping.json
```

Specify the following in `mapping.json`:

```json
[
  {
    "DeviceName": "/dev/sda1",
    "Ebs": {
      "DeleteOnTermination": false,
      "SnapshotId": "snap-1234567890abcdef0",
      "VolumeType": "gp2"
    }
  }
]
```

### Changing the Root Volume of a Running Instance to Persist Using the Command Line

You can use one of the following commands to change the root device volume of a running EBS-backed instance to persist. For more information about these command line interfaces, see [Accessing Amazon EC2](p. 3).

• **modify-instance-attribute** (AWS CLI)
• **Edit-EC2InstanceAttribute** (AWS Tools for Windows PowerShell)

For example, use the following command:

```
aws ec2 modify-instance-attribute --instance-id i-1234567890abcdef0 --block-device-mappings file://mapping.json
```

Specify the following in `mapping.json`:

```json
[
  {
    "DeviceName": "/dev/sda1",
    "Ebs": {
      "DeleteOnTermination": false
    }
  }
]
```

### Recover Your Instance

You can create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance and automatically recovers the instance if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair. Terminated instances cannot be recovered. A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata. If the impaired instance is in a placement group, the recovered instance runs in the placement group. For more information about using Amazon CloudWatch alarms to recover an instance,
see Create Alarms That Stop, Terminate, Reboot, or Recover an Instance (p. 565). To troubleshoot issues with instance recovery failures, see Troubleshooting Instance Recovery Failures (p. 394).

When the StatusCheckFailed_System alarm is triggered, and the recover action is initiated, you will be notified by the Amazon SNS topic that you selected when you created the alarm and associated the recover action. During instance recovery, the instance is migrated during an instance reboot, and any data that is in-memory is lost. When the process is complete, information is published to the SNS topic you've configured for the alarm. Anyone who is subscribed to this SNS topic will receive an email notification that includes the status of the recovery attempt and any further instructions. You will notice an instance reboot on the recovered instance.

Examples of problems that cause system status checks to fail include:

- Loss of network connectivity
- Loss of system power
- Software issues on the physical host
- Hardware issues on the physical host that impact network reachability

If your instance has a public IPv4 address, it retains the public IPv4 address after recovery.

Requirements

The recover action is supported only on instances with the following characteristics:

- One of the following instance types: C3, C4, C5, C5n, M3, M4, M5, M5a, P3, R3, R4, R5, R5a, T2, T3, T3a, X1, or X1e
- The default or dedicated instance tenancy
- EBS volumes only (do not configure instance store volumes)

Troubleshooting Instance Recovery Failures

The following issues can cause automatic recovery of your instance to fail:

- Temporary, insufficient capacity of replacement hardware.
- The instance has an attached instance store storage, which is an unsupported configuration for automatic instance recovery.
- There is an ongoing Service Health Dashboard event that prevented the recovery process from successfully executing. Refer to http://status.aws.amazon.com/ for the latest service availability information.
- The instance has reached the maximum daily allowance of three recovery attempts.

The automatic recovery process attempts to recover your instance for up to three separate failures per day. If the instance system status check failure persists, we recommend that you manually stop and start the instance. For more information, see Stop and Start Your Instance (p. 383).

Your instance may subsequently be retired if automatic recovery fails and a hardware degradation is determined to be the root cause for the original system status check failure.

Configuring Your Windows Instance

A Windows instance is a virtual server running Windows Server in the cloud.
After you have successfully launched and logged into your instance, you can make changes to it so that it's configured to meet the needs of a specific application. The following are some common tasks to help you get started.

Contents
- Configuring a Windows Instance Using EC2Launch (p. 395)
- Configuring a Windows Instance Using the EC2Config Service (p. 405)
- Paravirtual Drivers for Windows Instances (p. 431)
- AWS NVMe Drivers for Windows Instances (p. 446)
- Optimizing CPU Options (p. 447)
- Setting the Time for a Windows Instance (p. 460)
- Setting Passwords for Windows Instances (p. 464)
- Adding Windows Components Using Installation Media (p. 464)
- Configuring a Secondary Private IPv4 Address for Your Windows Instance (p. 468)
- Running Commands on Your Windows Instance at Launch (p. 472)
- Instance Metadata and User Data (p. 478)
- Best Practices and Recommendations for SQL Server Clustering in EC2 (p. 489)

Configuring a Windows Instance Using EC2Launch

EC2Launch is a set of Windows PowerShell scripts that replaces the EC2Config service on Windows Server 2016 and later AMIs.

Contents
- EC2Launch Tasks (p. 395)
- Installing the Latest Version of EC2Launch (p. 396)
- Verify the EC2Launch Version (p. 396)
- EC2Launch Directory Structure (p. 396)
- Configuring EC2Launch (p. 397)
- Using Sysprep with EC2Launch (p. 400)
- EC2Launch Version History (p. 403)

EC2Launch Tasks

EC2Launch performs the following tasks by default during the initial instance boot:

- Sets up new wallpaper that renders information about the instance.
- Sets the computer name.
- Sends instance information to the Amazon EC2 console.
- Sends the RDP certificate thumbprint to the EC2 console.
- Sets a random password for the administrator account.
- Adds DNS suffixes.
- Dynamically extends the operating system partition to include any unpartitioned space.
- Executes user data (if specified). For more information about specifying user data, see Working with Instance User Data (p. 480).
- Sets persistent static routes to reach the metadata service and KMS servers.
Important

If a custom AMI is created from this instance, these routes are captured as part of the OS configuration and any new instances launched from the AMI will retain the same routes, regardless of subnet placement. In order to update the routes, see Updating metadata/KMS routes for Server 2016 and later when launching a custom AMI (p. 402).

The following tasks help to maintain backward compatibility with the EC2Config service. You can also configure EC2Launch to perform these tasks during startup:

- Initialize secondary EBS volumes.
- Send Windows Event logs to the EC2 console logs.
- Send the *Windows is ready to use* message to the EC2 console.

For more information about Windows Server 2019, see Compare Features in Windows Server Versions on Microsoft.com.

Installing the Latest Version of EC2Launch

Use the following procedure to download and install the latest version of EC2Launch on your instances.

To download and install the latest version of EC2Launch

1. If you have already installed and configured EC2Launch on an instance, make a backup of the EC2Launch configuration file. The installation process does not preserve changes in this file. By default, the file is located in the `C:\ProgramData\Amazon\EC2-Windows\Launch\Config` directory.
2. Download `EC2-Windows-Launch.zip` to a directory on the instance.
3. Download `install.ps1` to the same directory where you downloaded `EC2-Windows-Launch.zip`.
4. Run `install.ps1`
5. If you made a backup of the EC2Launch configuration file, copy it to the `C:\ProgramData\Amazon\EC2-Windows\Launch\Config` directory.

Verify the EC2Launch Version

Use the following Windows PowerShell command to verify the installed version of EC2Launch.

```
PS C:\> Test-ModuleManifest -Path "C:\ProgramData\Amazon\EC2-Windows\Launch\Module\Ec2Launch.psd1" | Select Version
```

EC2Launch Directory Structure

EC2Launch is installed by default on Windows Server 2016 and later AMIs in the root directory `C:\ProgramData\Amazon\EC2-Windows\Launch`.

**Note**

By default, Windows hides files and folders under `C:\ProgramData`. To view EC2Launch directories and files, you must either type the path in Windows Explorer or change the folder properties to show hidden files and folders.

The Launch directory contains the following subdirectories.

- **Scripts** — Contains the PowerShell scripts that make up EC2Launch.
• Module — Contains the module for building scripts related to Amazon EC2.
• Config — Contains script configuration files that you can customize.
• Sysprep — Contains Sysprep resources.
• Settings — Contains an application for the Sysprep graphical user interface.
• Logs — Contains log files generated by scripts.

Configuring EC2Launch

After your instance has been initialized the first time, you can configure EC2Launch to run again and perform different start-up tasks.

Tasks
• Configure Initialization Tasks (p. 397)
• Schedule EC2Launch to Run on Every Boot (p. 398)
• Initialize Drives and Map Drive Letters (p. 398)
• Send Windows Event Logs to the EC2 Console (p. 399)
• Send Windows Is Ready Message After A Successful Boot (p. 399)

Configure Initialization Tasks

Specify settings in the LaunchConfig.json file to enable or disable the following initialization tasks:

• Set the computer name.
• Set up new wallpaper.
• Add DNS suffix list.
• Extend the boot volume size.
• Set the administrator password.

To configure initialization settings

1. On the instance to configure, open the following file in a text editor: C:\ProgramData\Amazon\EC2-Windows\Launch\Config\LaunchConfig.json.
2. Update the following settings as needed and save your changes. Provide a password in adminPassword only if adminPasswordType is Specify.

```json
{
    "setComputerName": false,
    "setWallpaper": true,
    "addDnsSuffixList": true,
    "extendBootVolumeSize": true,
    "handleUserData": true,
    "adminPasswordType": "Random | Specify | DoNothing",
    "adminPassword": "password that adheres to your security policy (optional)"
}
```

The password types are defined as follows:

Random

EC2Launch generates a password and encrypts it using the user's key. The system disables this setting after the instance is launched so that this password persists if the instance is rebooted or stopped and started.
Specify

EC2Launch uses the password you specify in `adminPassword`. If the password does not meet the system requirements, EC2Launch generates a random password instead. The password is stored in `LaunchConfig.json` as clear text and is deleted after Sysprep sets the administrator password. EC2Launch encrypts the password using the user's key.

DoNothing

EC2Launch uses the password you specify in the `unattend.xml` file. If you don't specify a password in `unattend.xml`, the administrator account is disabled.

3. In Windows PowerShell, run the following command to schedule the script to run as a Windows Scheduled Task. The script runs one time during the next boot and then disables these tasks from running again.

```
PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeInstance.ps1 - Schedule
```

Schedule EC2Launch to Run on Every Boot

You can schedule EC2Launch to run on every boot instead of only the initial boot.

To enable EC2Launch to run on every boot:

1. Open Windows PowerShell and run the following command:

```
PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeInstance.ps1 - SchedulePerBoot
```

2. Or, run the executable with the following command:

```
PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Settings\Ec2LaunchSettings.exe
```

Then select Run EC2Launch on every boot. You can specify that your EC2 instance Shutdown without Sysprep or Shutdown with Sysprep.

Note

When you enable EC2Launch to run on every boot, the following changes will be made to the `LaunchConfig.json` the next time EC2Launch runs:

- `AdminPasswordType` will be set back to DoNothing so that the password does not change on each boot.
- `HandleUserData` will be set back to false unless the user data has persist set to true. For more information about user data scripts, see User Data Scripts in the Amazon EC2 User Guide.

Similarly, if you do not want your password reset on the next boot, you should set `AdminPasswordType` to DoNothing before rebooting.

Initialize Drives and Map Drive Letters

Specify settings in the `DriveLetterMappingConfig.json` file to map drive letters to volumes on your EC2 instance. The script performs this operation if the drives have not already been initialized and partitioned.
To map drive letters to volumes

1. Open the C:\ProgramData\Amazon\EC2-Windows\Launch\Config \DriveLetterMappingConfig.json file in a text editor.
2. Specify the following volume settings and save your changes:

   ```json
   {
     "driveLetterMapping": [
       {
         "volumeName": "sample volume",
         "driveLetter": "H"
       }
     ]
   }
   ```

3. Open Windows PowerShell and use the following command to run the EC2Launch script that initializes the disks:

   ```powershell
   PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeDisks.ps1
   ```

To initialize the disks each time the instance boots, add the `-Schedule` flag as follows:

```powershell
PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeDisks.ps1 -Schedule
```

Send Windows Event Logs to the EC2 Console

Specify settings in the EventLogConfig.json file to send Windows Event logs to EC2 console logs.

To configure settings to send Windows Event logs

1. On the instance, open the C:\ProgramData\Amazon\EC2-Windows\Launch\Config \EventLogConfig.json file in a text editor.
2. Configure the following log settings and save your changes:

   ```json
   {
     "events": [
       {
         "logName": "System",
         "source": "An event source (optional)",
         "level": "Error | Warning | Information",
         "numEntries": 3
       }
     ]
   }
   ```

3. In Windows PowerShell, run the following command so that the system schedules the script to run as a Windows Scheduled Task each time the instance boots.

   ```powershell
   PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\SendEventLogs.ps1 -Schedule
   ```

The logs can take three minutes or more to appear in the EC2 console logs.

Send Windows Is Ready Message After A Successful Boot

The EC2Config service sent the "Windows is ready" message to the EC2 console after every boot. EC2Launch sends this message only after the initial boot. For backwards compatibility with the
EC2Config service, you can schedule EC2Launch to send this message after every boot. On the instance, open Windows PowerShell and run the following command. The system schedules the script to run as a Windows Scheduled Task.

```
PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\SendWindowsIsReady.ps1 -Schedule
```

**Using Sysprep with EC2Launch**

Sysprep simplifies the process of duplicating a customized installation of Windows Server 2016 and later. EC2Launch offers a default answer file and batch files for Sysprep that automate and secure the image-preparation process on your AMI. Modifying these files is optional. These files are located in the following directory by default: `C:\ProgramData\Amazon\EC2-Windows\Launch\Sysprep`.

**Important**

Do not use Sysprep to create an instance backup. Sysprep removes system-specific information. If you remove this information there might be unintended consequences for an instance backup.

The EC2Launch answer file and batch files for Sysprep include the following:

- **Unattend.xml**
  
  This is the default answer file. If you run `SysprepInstance.ps1` or choose `ShutdownWithSysprep` in the user interface, the system reads the setting from this file.

- **BeforeSysprep.cmd**

  Customize this batch file to run commands before EC2Launch runs Sysprep.

- **SysprepSpecialize.cmd**

  Customize this batch file to run commands during the Sysprep specialize phase.

**Running Sysprep with EC2Launch**

On the full installation of Windows Server 2016 and later (with a desktop experience), you can run Sysprep with EC2Launch manually or by using the **EC2 Launch Settings** application.

**To run Sysprep using the EC2Launch Settings application**

1. In the Amazon EC2 console, locate or create a Windows Server 2016 or later AMI.
2. Launch a Windows instance from the AMI.
3. Connect to your Windows instance and customize it.
4. Search for and run the **EC2LaunchSettings** application. It is located in the following directory by default: `C:\ProgramData\Amazon\EC2-Windows\Launch\Settings`. 
5. Select or clear options as needed. These settings are stored in the `LaunchConfig.json` file.

6. For **Administrator Password**, do one of the following:
   - Choose **Random**. EC2Launch generates a password and encrypts it using the user's key. The system disables this setting after the instance is launched so that this password persists if the instance is rebooted or stopped and started.
   - Choose **Specify** and type a password that meets the system requirements. The password is stored in `LaunchConfig.json` as clear text and is deleted after Sysprep sets the administrator password. If you shut down now, the password is set immediately. EC2Launch encrypts the password using the user's key.
   - Choose **DoNothing** and specify a password in the `unattend.xml` file. If you don't specify a password in `unattend.xml`, the administrator account is disabled.

7. Choose **Shutdown with Sysprep**.

**To manually run Sysprep using EC2Launch**

1. In the Amazon EC2 console locate or create a Windows Server 2016 or later Datacenter edition AMI that you want to duplicate.
2. Launch and connect to your Windows instance.
3. Customize the instance.
4. Specify settings in the LaunchConfig.json file. This file is located in the C:\ProgramData\Amazon\EC2-Windows\Launch\Config directory by default.

For adminPasswordType, specify one of the following values:

Random

EC2Launch generates a password and encrypts it using the user's key. The system disables this setting after the instance is launched so that this password persists if the instance is rebooted or stopped and started.

Specify

EC2Launch uses the password you specify in adminPassword. If the password does not meet the system requirements, EC2Launch generates a random password instead. The password is stored in LaunchConfig.json as clear text and is deleted after Sysprep sets the administrator password. EC2Launch encrypts the password using the user's key.

DoNothing

EC2Launch uses the password you specify in the unattend.xml file. If you don't specify a password in unattend.xml, the administrator account is disabled.

5. (Optional) Specify settings in unattend.xml and other configuration files. If plan to attend to the installation, then you don't need to make changes in these files. The files are located in the following directory by default: C:\ProgramData\Amazon\EC2-Windows\Launch\Sysprep.

6. In Windows PowerShell, run ./InitializeInstance.ps1 -Schedule. The script is located in the following directory by default: C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts. This script schedules the instance to initialize during the next boot. You must run this script before you run the SysprepInstance.ps1 script in the next step.

7. In Windows PowerShell, run ./SysprepInstance.ps1. The script is located in the following directory by default: C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts.

You are logged off the instance and the instance shuts down. If you check the Instances page in the Amazon EC2 console, the instance state changes from running to stopping, and then to stopped. At this point, it is safe to create an AMI from this instance.

**Updating metadata/KMS routes for Server 2016 and later when launching a custom AMI**

To update metadata/KMS routes for Server 2016 and later when launching a custom AMI

1. Use EC2LaunchSettings GUI (C:\ProgramData\Amazon\EC2-Windows\Launch\Settings\Ec2LaunchSettings.exe) to shut down with Sysprep.
2. Or, shut down without Sysprep before creating an AMI. This sets the EC2 Launch Initialize tasks to run at the next boot, which will set routes based on the subnet being launched into.
3. Or, manually reschedule EC2 Launch initialize tasks before creating an AMI from PowerShell (p. 397).

**Important**

Please take note of the default password reset behavior before rescheduling tasks.
EC2Launch Version History

Windows AMIs starting with Windows Server 2016 include a set of Windows Powershell scripts called EC2Launch. EC2Launch performs tasks during the initial instance boot. For information about the EC2Launch versions included in the Windows AMIs, see see [Managed AWS Windows AMIs](p. 78).

To download and install the latest version of EC2Launch, see [Installing the Latest Version of EC2Launch](p. 396).

The following table describes the released versions of EC2Launch.

<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3.2001660</td>
<td>• Fixed automatic login issue for users without password after first time executing Sysprep.</td>
</tr>
<tr>
<td>1.3.2001360</td>
<td>• Fixed minor issues.</td>
</tr>
<tr>
<td>1.3.2001220</td>
<td>• All PowerShell scripts signed.</td>
</tr>
<tr>
<td>1.3.2001200</td>
<td>• Fixed issue with InitializeDisks.ps1 where running the script on a node in a Microsoft Windows Server Failover Cluster would format drives on remote nodes whose drive letter matched the local drive letter.</td>
</tr>
<tr>
<td>1.3.2001160</td>
<td>• Fixed missing wallpaper in Windows 2019.</td>
</tr>
<tr>
<td>1.3.2001040</td>
<td>• Added plugin for setting the monitor to never turn off to fix ACPI issues.</td>
</tr>
<tr>
<td></td>
<td>• SQL Server edition and version written to console.</td>
</tr>
<tr>
<td>1.3.2000930</td>
<td>• Fix for adding routes to metadata on ipv6-enabled ENIs.</td>
</tr>
<tr>
<td>1.3.2000760</td>
<td>• Added default configuration for RSS and Receive Queue settings for ENA devices.</td>
</tr>
<tr>
<td></td>
<td>• Disabled hibernation during Sysprep.</td>
</tr>
<tr>
<td>1.3.2000630</td>
<td>• Added route 169.254.169.253/32 for DNS server.</td>
</tr>
<tr>
<td></td>
<td>• Added filter of setting Admin user.</td>
</tr>
<tr>
<td></td>
<td>• Improvements made to instance hibernation.</td>
</tr>
<tr>
<td></td>
<td>• Added option to schedule EC2Launch to run on every boot.</td>
</tr>
<tr>
<td>1.3.2000430.0</td>
<td>• Added route 169.254.169.123/32 to AMZN time service.</td>
</tr>
<tr>
<td></td>
<td>• Added route 169.254.169.249/32 to GRID license service.</td>
</tr>
<tr>
<td></td>
<td>• Added timeout of 25 seconds when attempting to start Systems Manager.</td>
</tr>
<tr>
<td>1.3.200039.0</td>
<td>• Fixed improper drive lettering for EBS NVME volumes.</td>
</tr>
<tr>
<td>Version</td>
<td>Details</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.3.2000080</td>
<td>Fixed minor issues.</td>
</tr>
<tr>
<td>1.3.610</td>
<td>Fixed issue with redirecting output and errors to files from user data.</td>
</tr>
<tr>
<td>1.3.590</td>
<td>• Added missing instances types in the wallpaper.</td>
</tr>
<tr>
<td></td>
<td>• Fixed an issue with drive letter mapping and disk installation.</td>
</tr>
<tr>
<td>1.3.580</td>
<td>• Fixed Get-Metadata to use the default system proxy settings for web requests.</td>
</tr>
<tr>
<td></td>
<td>• Added a special case for NVMe in disk initialization.</td>
</tr>
<tr>
<td></td>
<td>• Fixed minor issues.</td>
</tr>
<tr>
<td>1.3.550</td>
<td>Added a -NoShutdown option to enable Sysprep with no shutdown.</td>
</tr>
<tr>
<td>1.3.540</td>
<td>Fixed minor issues.</td>
</tr>
<tr>
<td>1.3.530</td>
<td>Fixed minor issues.</td>
</tr>
<tr>
<td>1.3.521</td>
<td>Fixed minor issues.</td>
</tr>
<tr>
<td>1.3.0</td>
<td>• Fixed a hexadecimal length issue for computer name change.</td>
</tr>
<tr>
<td></td>
<td>• Fixed a possible reboot loop for computer name change.</td>
</tr>
<tr>
<td></td>
<td>• Fixed an issue in wallpaper setup.</td>
</tr>
<tr>
<td>1.2.0</td>
<td>• Update to display information about installed operating system (OS) in EC2 system log.</td>
</tr>
<tr>
<td></td>
<td>• Update to display EC2Launch and SSM Agent version in EC2 system log.</td>
</tr>
<tr>
<td></td>
<td>• Fixed minor issues.</td>
</tr>
<tr>
<td>1.1.2</td>
<td>• Update to display ENA driver information in EC2 system log.</td>
</tr>
<tr>
<td></td>
<td>• Update to exclude Hyper-V from primary NIC filter logic.</td>
</tr>
<tr>
<td></td>
<td>• Added KMS server and port into registry key for KMS activation.</td>
</tr>
<tr>
<td></td>
<td>• Improved wallpaper setup for multiple users.</td>
</tr>
<tr>
<td></td>
<td>• Update to clear routes from persistent store.</td>
</tr>
<tr>
<td></td>
<td>• Update to remove the z from availability zone in DNS suffix list.</td>
</tr>
<tr>
<td></td>
<td>• Update to address an issue with the &lt;runAsLocalSystem&gt; tag in user data.</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Initial release.</td>
</tr>
</tbody>
</table>
Configuring a Windows Instance Using the EC2Config Service

Windows AMIs for Windows Server 2012 R2 and earlier include an optional service, the EC2Config service (EC2Config.exe). EC2Config starts when the instance boots and performs tasks during startup and each time you stop or start the instance. EC2Config can also perform tasks on demand. Some of these tasks are automatically enabled, while others must be enabled manually. Although optional, this service provides access to advanced features that aren't otherwise available. This service runs in the LocalSystem account.

Note
EC2Launch replaces EC2Config on Windows AMIs for Windows Server 2016 and later. For more information, see Configuring a Windows Instance Using EC2Launch (p. 395).

EC2Config uses settings files to control its operation. You can update these settings files using either a graphical tool or by directly editing XML files. The service binaries and additional files are contained in the %ProgramFiles%\Amazon\EC2ConfigService directory.

Contents
• EC2Config Tasks (p. 405)
• Installing the Latest Version of EC2Config (p. 406)
• Stopping, Restarting, Deleting, or Uninstalling EC2Config (p. 407)
• EC2Config and AWS Systems Manager (p. 408)
• EC2Config and Sysprep (p. 408)
• EC2 Service Properties (p. 408)
• EC2Config Settings Files (p. 411)
• Configure Proxy Settings for the EC2Config Service (p. 415)
• EC2Config Version History (p. 417)
• Troubleshooting Issues with the EC2Config Service (p. 429)

EC2Config Tasks

EC2Config runs initial startup tasks when the instance is first started and then disables them. To run these tasks again, you must explicitly enable them prior to shutting down the instance, or by running Sysprep manually. These tasks are as follows:

• Set a random, encrypted password for the administrator account.
• Generate and install the host certificate used for Remote Desktop Connection.
• Dynamically extend the operating system partition to include any unpartitioned space.
• Execute the specified user data (and Cloud-Init, if it's installed). For more information about specifying user data, see Working with Instance User Data (p. 480).

EC2Config performs the following tasks every time the instance starts:

• Change the host name to match the private IP address in Hex notation (this task is disabled by default and must be enabled in order to run at instance start).
• Configure the key management server (AWS KMS), check for Windows activation status, and activate Windows as necessary.
• Mount all Amazon EBS volumes and instance store volumes, and map volume names to drive letters.
• Write event log entries to the console to help with troubleshooting (this task is disabled by default and must be enabled in order to run at instance start).
• Write to the console that Windows is ready.
• Add a custom route to the primary network adapter to enable the following IP addresses when multiple NICs are attached: 169.254.169.250, 169.254.169.251, and 169.254.169.254. These addresses are used by Windows Activation and when you access instance metadata.

EC2Config performs the following task every time a user logs in:
• Display wallpaper information to the desktop background.

While the instance is running, you can request that EC2Config perform the following task on demand:
• Run Sysprep and shut down the instance so that you can create an AMI from it. For more information, see Create a Standard Amazon Machine Image Using Sysprep (p. 113).

Installing the Latest Version of EC2Config

By default, the EC2Config service is included in AMIs prior to Windows Server 2016. When the EC2Config service is updated, new Windows AMIs from AWS include the latest version of the service. However, you need to update your own Windows AMIs and instances with the latest version of EC2Config.

Note
EC2Launch replaces EC2Config on Windows Server 2016 and later AMIs. For more information, see Configuring a Windows Instance Using EC2Launch (p. 395).

For information about how to receive notifications for EC2Config updates, see Subscribing to EC2Config Service Notifications (p. 429). For information about the changes in each version, see the EC2Config Version History (p. 417).

Before You Begin
• Verify that you have .NET framework 3.5 SP1 or greater.
• By default, Setup replaces your settings files with default settings files during installation and restarts the EC2Config service when the installation is completed. If you changed EC2Config service settings, copy the config.xml file from the %Program Files%\Amazon\Ec2ConfigService\Settings directory. After you update the EC2Config service, you can restore this file to retain your configuration changes.
• If your version of EC2Config is earlier than version 2.1.19 and you are installing version 2.2.12 or earlier, you must first install version 2.1.19. To install version 2.1.19, download EC2Install_2.1.19.zip, unzip the file, and then run EC2Install.exe.

Note
If your version of EC2Config is earlier than version 2.1.19 and you are installing version 2.3.313 or later, you can install it directly without installing version 2.1.19 first.

Verify the EC2Config Version

Use the following procedure to verify the version of EC2Config that is installed on your instances.

To verify the installed version of EC2Config
1. Launch an instance from your AMI and connect to it.
2. In Control Panel, select Programs and Features.
3. In the list of installed programs, look for Ec2ConfigService. Its version number appears in the Version column.

**Update EC2Config**

Use the following procedure to download and install the latest version of EC2Config on your instances.

**To download and install the latest version of EC2Config**

1. Download and unzip the EC2Config installer.
2. Run EC2Install.exe. For a complete list of options, run EC2Install with the /? option. By default, setup displays prompts. To run the command with no prompts, use the /quiet option.
   
   **Important**
   
   To keep the custom settings from the config.xml file that you saved, run EC2Install with the /norestart option, restore your settings, and then restart the EC2Config service manually.
3. If you are running EC2Config version 4.0 or later, you must restart SSM Agent on the instance from the Microsoft Services snap-in.

   **Note**
   
   The updated EC2Config version information will not appear in the instance System Log or Trusted Advisor check until you reboot or stop and start your instance.

**Stopping, Restarting, Deleting, or Uninstalling EC2Config**

You can manage the EC2Config service just as you would any other service.

To apply updated settings to your instance, you can stop and restart the service. If you're manually installing EC2Config, you must stop the service first.

**To stop the EC2Config service**

1. Launch and connect to your Windows instance.
2. On the Start menu, point to Administrative Tools, and then click Services.
3. In the list of services, right-click EC2Config, and select Stop.

**To restart the EC2Config service**

1. Launch and connect to your Windows instance.
2. On the Start menu, point to Administrative Tools, and then click Services.
3. In the list of services, right-click EC2Config, and select Restart.

If you don’t need to update the configuration settings, create your own AMI, or use AWS Systems Manager, you can delete and uninstall the service. Deleting a service removes its registry subkey. Uninstalling a service removes the files, the registry subkey, and any shortcuts to the service.

**To delete the EC2Config service**

1. Start a command prompt window.
2. Run the following command:

   ```bash
   sc delete ec2config
   ```
To uninstall EC2Config

1. Launch and connect to your Windows instance.
2. On the Start menu, click Control Panel.
3. Double-click Programs and Features.
4. On the list of programs, select EC2ConfigService, and click Uninstall.

EC2Config and AWS Systems Manager

The EC2Config service processes Systems Manager requests on instances created from AMIs for versions of Windows Server prior to Windows Server 2016 that were published before November 2016.

Instances created from AMIs for versions of Windows Server prior to Windows Server 2016 that were published after November 2016 include the EC2Config service and SSM Agent. EC2Config performs all of the tasks described earlier, and SSM Agent processes requests for Systems Manager capabilities like Run Command and State Manager.

You can use Run Command to upgrade your existing instances to use to the latest version of the EC2Config service and SSM Agent. For more information, see Update SSM Agent by using Run Command in the AWS Systems Manager User Guide.

EC2Config and Sysprep

The EC2Config service runs Sysprep, a Microsoft tool that enables you to create a customized Windows AMI that can be reused. When EC2Config calls Sysprep, it uses the files in %ProgramFiles%\Amazon\EC2ConfigService\Settings to determine which operations to perform. You can edit these files indirectly using the Ec2 Service Properties dialog box, or directly using an XML editor or a text editor. However, there are some advanced settings that aren't available in the Ec2 Service Properties dialog box, so you must edit those entries directly.

If you create an AMI from an instance after updating its settings, the new settings are applied to any instance that's launched from the new AMI. For information about creating an AMI, see Creating a Custom Windows AMI (p. 64).

EC2 Service Properties

The following procedure describes how to use the Ec2 Service Properties dialog box to enable or disable settings.

To change settings using the Ec2 Service Properties dialog box

1. Launch and connect to your Windows instance.
2. From the Start menu, click All Programs, and then click EC2ConfigService Settings.
3. On the General tab of the Ec2 Service Properties dialog box, you can enable or disable the following settings.

Set Computer Name

If this setting is enabled (it is disabled by default), the host name is compared to the current internal IP address at each boot; if the host name and internal IP address do not match, the host name is reset to contain the internal IP address and then the system reboots to pick up the new host name. To set your own host name, or to prevent your existing host name from being modified, do not enable this setting.

User Data

User data execution enables you to specify scripts in the instance metadata. By default, these scripts are run during the initial launch. You can also configure them to run the next time you reboot or start the instance, or every time you reboot or start the instance.

If you have a large script, we recommend that you use user data to download the script, and then execute it.

For more information, see User Data Execution (p. 473).

Event Log

Use this setting to display event log entries on the console during boot for easy monitoring and debugging.

Click Settings to specify filters for the log entries sent to the console. The default filter sends the three most recent error entries from the system event log to the console.

Wallpaper Information

Use this setting to display system information on the desktop background. The following is an example of the information displayed on the desktop background.
The information displayed on the desktop background is controlled by the settings file EC2ConfigService\Settings\WallpaperSettings.xml.

Enable Hibernation

Use this setting to allow EC2 to signal the operating system to perform hibernation.

4. Click the Storage tab. You can enable or disable the following settings.

Root Volume

This setting dynamically extends Disk 0/Volume 0 to include any unpartitioned space. This can be useful when the instance is booted from a root device volume that has a custom size.

Initialize Drives

This setting formats and mounts all volumes attached to the instance during start.

Drive Letter Mapping

The system maps the volumes attached to an instance to drive letters. For Amazon EBS volumes, the default is to assign drive letters going from D: to Z:. For instance store volumes, the default depends on the driver. Citrix PV drivers assign instance store volumes drive letters going from Z: to A:. Red Hat drivers assign instance store volumes drive letters going from D: to Z:.

To choose the drive letters for your volumes, click Mappings. In the DriveLetterSetting dialog box, specify the Volume Name and Drive Letter values for each volume, click Apply, and then
click OK. We recommend that you select drive letters that avoid conflicts with drive letters that are likely to be in use, such as drive letters in the middle of the alphabet.

![DriveLetterSetting](image)

After you specify a drive letter mapping and attach a volume with same label as one of the volume names that you specified, EC2Config automatically assigns your specified drive letter to that volume. However, the drive letter mapping fails if the drive letter is already in use. Note that EC2Config doesn't change the drive letters of volumes that were already mounted when you specified the drive letter mapping.

5. To save your settings and continue working on them later, click OK to close the Ec2 Service Properties dialog box. If you have finished customizing your instance and want to create an AMI from that instance, see Create a Standard Amazon Machine Image Using Sysprep (p. 113).

**EC2Config Settings Files**

The settings files control the operation of the EC2Config service. These files are located in the `C:\Program Files\Amazon\Ec2ConfigService\Settings` directory:

- `ActivationSettings.xml`—Controls product activation using a key management server (KMS).
- `AWS.EC2.Windows.CloudWatch.json`—Controls which performance counters to send to CloudWatch and which logs to send to CloudWatch Logs.
- `BundleConfig.xml`—Controls how EC2Config prepares an instance store-backed instance for AMI creation.
- `Config.xml`—Controls the primary settings.
- `DriveLetterConfig.xml`—Controls drive letter mappings.
- `EventLogConfig.xml`—Controls the event log information that's displayed on the console while the instance is booting.
- `WallpaperSettings.xml`—Controls the information that's displayed on the desktop background.

**ActivationSettings.xml**

This file contains settings that control product activation. When Windows boots, the EC2Config service checks whether Windows is already activated. If Windows is not already activated, it attempts to activate Windows by searching for the specified KMS server.

- `SetAutodiscover`—Indicates whether to detect a KMS automatically.
- `TargetKMSserver`—Stores the private IP address of a KMS. The KMS must be in the same region as your instance.
- `DiscoverFromZone`—Discovers the KMS server from the specified DNS zone.
- `ReadFromUserData`—Gets the KMS server from UserData.
- `LegacySearchZones`—Discovers the KMS server from the specified DNS zone.
- `DoActivate`—Attempts activation using the specified settings in the section. This value can be `true` or `false`. 

---

411
• LogResultToConsole—Displays the result to the console.

**BundleConfig.xml**

This file contains settings that control how EC2Config prepares an instance for AMI creation.

• AutoSysprep—Indicates whether to use Sysprep automatically. Change the value to Yes to use Sysprep.

• SetRDPCertificate—Sets a self-signed certificate to the Remote Desktop server. This enables you to securely RDP into the instances. Change the value to Yes if the new instances should have the certificate.

  This setting is not used with Windows Server 2008 or Windows Server 2012 instances because they can generate their own certificates.

• SetPasswordAfterSysprep—Sets a random password on a newly launched instance, encrypts it with the user launch key, and outputs the encrypted password to the console. Change the value of this setting to No if the new instances should not be set to a random encrypted password.

**Config.xml**

**Plug-ins**

• Ec2SetPassword—Generates a random encrypted password each time you launch an instance. This feature is disabled by default after the first launch so that reboots of this instance don't change a password set by the user. Change this setting to Enabled to continue to generate passwords each time you launch an instance.

  This setting is important if you are planning to create an AMI from your instance.

• Ec2SetComputerName—Sets the host name of the instance to a unique name based on the IP address of the instance and reboots the instance. To set your own host name, or prevent your existing host name from being modified, you must disable this setting.

• Ec2InitializeDrives—Initializes and formats all volumes during startup. This feature is enabled by default.

• Ec2EventLog—Displays event log entries in the console. By default, the three most recent error entries from the system event log are displayed. To specify the event log entries to display, edit the EventLogConfig.xml file located in the EC2ConfigService\Settings directory. For information about the settings in this file, see Eventlog Key in the MSDN Library.

• Ec2ConfigureRDP—Sets up a self-signed certificate on the instance, so users can securely access the instance using Remote Desktop. This feature is disabled on Windows Server 2008 and Windows Server 2012 instances because they can generate their own certificates.

• Ec2OutputRDPert—Displays the Remote Desktop certificate information to the console so that the user can verify it against the thumbprint.

• Ec2SetDriveLetter—Sets the drive letters of the mounted volumes based on user-defined settings. By default, when an Amazon EBS volume is attached to an instance, it can be mounted using the drive letter on the instance. To specify your drive letter mappings, edit the DriveLetterConfig.xml file located in the EC2ConfigService\Settings directory.

• Ec2WindowsActivate—The plug-in handles Windows activation. It checks to see if Windows is activated. If not, it updates the KMS client settings, and then activates Windows.

  To modify the KMS settings, edit the ActivationSettings.xml file located in the EC2ConfigService\Settings directory.

• Ec2DynamicBootVolumeSize—Extends Disk 0/Volume 0 to include any unpartitioned space.

• Ec2HandleUserData—Creates and executes scripts created by the user on the first launch of an instance after Sysprep is run. Commands wrapped in script tags are saved to a batch file, and
commands wrapped in PowerShell tags are saved to a .ps1 file (corresponds to the User Data check box on the Ec2 Service Properties dialog box).

- **Ec2ElasticGpuSetup**—Installs the Elastic GPU software package if the instance is associated with an elastic GPU.
- **Ec2FeatureLogging**—Sends Windows feature installation and corresponding service status to the console. Supported only for the Microsoft Hyper-V feature and corresponding vmms service.

**Global Settings**

- **ManageShutdown**—Ensures that instances launched from instance store-backed AMIs do not terminate while running Sysprep.
- **SetDnsSuffixList**—Sets the DNS suffix of the network adapter for Amazon EC2. This allows DNS resolution of servers running in Amazon EC2 without providing the fully qualified domain name.
- **WaitForMetaDataAvailable**—Ensures that the EC2Config service will wait for metadata to be accessible and the network available before continuing with the boot. This check ensures that EC2Config can obtain information from metadata for activation and other plug-ins.
- **ShouldAddRoutes**—Adds a custom route to the primary network adapter to enable the following IP addresses when multiple NICs are attached: 169.254.169.250, 169.254.169.251, and 169.254.169.254. These addresses are used by Windows Activation and when you access instance metadata.
- **RemoveCredentialsfromSyspreponStartup**—Removes the administrator password from Sysprep.xml the next time the service starts. To ensure that this password persists, edit this setting.

**DriveLetterConfig.xml**

This file contains settings that control drive letter mappings. By default, a volume can be mapped to any available drive letter. You can mount a volume to a particular drive letter as follows.

```
<?xml version="1.0" standalone="yes"?>
<DriveLetterMapping>
  <Mapping>
    <VolumeName></VolumeName>
    <DriveLetter></DriveLetter>
  </Mapping>
  . . .
  <Mapping>
    <VolumeName></VolumeName>
    <DriveLetter></DriveLetter>
  </Mapping>
</DriveLetterMapping>
```

- **VolumeName**—The volume label. For example, My Volume. To specify a mapping for an instance storage volume, use the label Temporary Storage X, where X is a number from 0 to 25.
- **DriveLetter**—The drive letter. For example, M:. The mapping fails if the drive letter is already in use.

**EventLogConfig.xml**

This file contains settings that control the event log information that's displayed on the console while the instance is booting. By default, we display the three most recent error entries from the System event log.

- **Category**—The event log key to monitor.
- **ErrorType**—The event type (for example, Error, Warning, Information.)
- **NumEntries**—The number of events stored for this category.
• LastMessageTime—To prevent the same message from being pushed repeatedly, the service updates this value every time it pushes a message.
• AppName—The event source or application that logged the event.

WallpaperSettings.xml

This file contains settings that control the information that’s displayed on the desktop background. The following information is displayed by default.

• Hostname—Displays the computer name.
• Instance ID—Displays the ID of the instance.
• Public IP Address—Displays the public IP address of the instance.
• Private IP Address—Displays the private IP address of the instance.
• Availability Zone—Displays the Availability Zone in which the instance is running.
• Instance Size—Displays the type of instance.
• Architecture—Displays the setting of the PROCESSOR_ARCHITECTURE environment variable.

You can remove any of the information that’s displayed by default by deleting its entry. You can add additional instance metadata to display as follows.

```xml
<WallpaperInformation>
  <name>display_name</name>
  <source>metadata</source>
  <identifier>meta-data/path</identifier>
</WallpaperInformation>
```

You can add additional System environment variables to display as follows.

```xml
<WallpaperInformation>
  <name>display_name</name>
  <source>EnvironmentVariable</source>
  <identifier>variable-name</identifier>
</WallpaperInformation>
```

InitializeDrivesSettings.xml

This file contains settings that control how EC2Config initializes drives.

By default, EC2Config initialize drives that were not brought online with the operating system. You can customize the plugin as follows.

```xml
<InitializeDrivesSettings>
  <SettingsGroup>setting</SettingsGroup>
</InitializeDrivesSettings>
```

Use a settings group to specify how you want to initialize drives:

**FormatWithTRIM**

Enables the TRIM command when formatting drives. After a drive has been formatted and initialized, the system restores TRIM configuration.

Starting with EC2Config version 3.18, the TRIM command is disabled during the disk format operation by default. This improves formatting times. Use this setting to enable TRIM during the disk format operation for EC2Config version 3.18 and later.
FormatWithoutTRIM

Disables the TRIM command when formatting drives and improves formatting times in Windows. After a drive has been formatted and initialized, the system restores TRIM configuration.

DisableInitializeDrives

Disables formatting for new drives. Use this setting to initialize drives manually.

Configure Proxy Settings for the EC2Config Service

You can configure the EC2Config service to communicate through a proxy using one of the following methods: the AWS SDK for .NET, the system.net element, or Microsoft Group Policy and Internet Explorer. Using the AWS SDK for .NET is the preferred method because you can specify a user name and password.

Methods

- Configure Proxy Settings Using the AWS SDK for .NET (Preferred) (p. 415)
- Configure Proxy Settings Using the system.net Element (p. 416)
- Configure Proxy Settings Using Microsoft Group Policy and Microsoft Internet Explorer (p. 416)

Configure Proxy Settings Using the AWS SDK for .NET (Preferred)

You can configure proxy settings for the EC2Config service by specifying the proxy element in the Ec2Config.exe.config file. For more information, see Configuration Files Reference for AWS SDK for .NET.

To specify the proxy element in Ec2Config.exe.config

1. Edit the Ec2Config.exe.config file on an instance where you want the EC2Config service to communicate through a proxy. By default, the file is located in the following directory: %ProgramFiles%\Amazon\Ec2ConfigService.
2. Add the following aws element to the configSections. Do not add this to any existing sectionGroups.

For EC2Config versions 3.17 or earlier

```xml
<configSections>
  <section name="aws" type="Amazon.AWSSection, AWSSDK"/>
</configSections>
```

For EC2Config versions 3.18 or later

```xml
<configSections>
  <section name="aws" type="Amazon.AWSSection, AWSSDK.Core"/>
</configSections>
```

3. Add the following aws element to the Ec2Config.exe.config file.

```xml
<aws>
  <proxy
    host="string value"
    port="string value"
    username="string value"
    password="string value" />
```

415
4. Save your changes.

Configure Proxy Settings Using the system.net Element

You can specify proxy settings in a system.net element in the Ec2Config.exe.config file. For more information, see defaultProxy Element (Network Settings) on MSDN.

To specify the system.net element in Ec2Config.exe.config

1. Edit the Ec2Config.exe.config file on an instance where you want the EC2Config service to communicate through a proxy. By default, the file is located in the following directory: %ProgramFiles%\Amazon\Ec2ConfigService.
2. Add a defaultProxy entry to system.net. For more information, see defaultProxy Element (Network Settings) on MSDN.

For example, the following configuration routes all traffic to use the proxy that is currently configured for Internet Explorer, with the exception of the metadata and licensing traffic, which will bypass the proxy.

```xml
<defaultProxy>
  <proxy usesystemdefault="true" />
  <bypasslist>
    <add address="169.254.169.250" />
    <add address="169.254.169.251" />
    <add address="169.254.169.254" />
  </bypasslist>
</defaultProxy>
```

3. Save your changes.

Configure Proxy Settings Using Microsoft Group Policy and Microsoft Internet Explorer

The EC2Config service runs under the Local System user account. You can specify instance-wide proxy settings for this account in Internet Explorer after you change Group Policy settings on the instance.

To configure proxy settings using Group Policy and Internet Explorer

1. On an instance where you want the EC2Config service to communicate through a proxy, open a Command prompt as an Administrator, type `gpedit.msc`, and press Enter.
2. In the Local Group Policy Editor, under Local Computer Policy, choose Computer Configuration, Administrative Templates, Windows Components, Internet Explorer.
3. In the right-pane, choose Make proxy settings per-machine (rather than per-user) and then choose Edit policy setting.
4. Choose Enabled, and then choose Apply.
5. Open Internet Explorer, and then choose the Tools button.
6. Choose Internet Option, and then choose the Connections tab.
7. Choose LAN settings.
8. Under Proxy server, choose the Use a proxy server for your LAN option.
9. Specify address and port information and then choose OK.
EC2Config Version History

Windows AMIs prior to Windows Server 2016 include an optional service called the EC2Config service (EC2Config.exe). EC2Config starts when the instance boots and performs tasks during startup and each time you stop or start the instance. For information about the EC2Config versions included in the Windows AMIs, see Managed AWS Windows AMIs (p. 78).

You can receive notifications when new versions of the EC2Config service are released. For more information, see Subscribing to EC2Config Service Notifications (p. 429).

The following table describes the released versions of EC2Config. For information about the updates for SSM Agent, see Systems Manager SSM Agent Release Notes.

<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9.3519</td>
<td>• New version of SSM Agent 2.3.634.0</td>
</tr>
<tr>
<td>4.9.3429</td>
<td>• New version of SSM Agent 2.3.542.0</td>
</tr>
<tr>
<td>4.9.3289</td>
<td>• New version of SSM Agent 2.3.444.0</td>
</tr>
</tbody>
</table>
| 4.9.3270 | • Added plugin for setting the monitor to never turn off to fix ACPI issues  
| | • SQL Server edition and version written to console  
| | • New version of SSM Agent 2.3.415.0 |
| 4.9.3230 | • Drive Letter Mapping description updated to better align to functionality  
| | • New version of SSM Agent 2.3.372.0 |
| 4.9.3160 | • Increased wait time for primary NIC  
| | • Added default configuration for RSS and Receive Queue settings for ENA devices  
| | • Disabled hibernation during Sysprep  
| | • New version of SSM Agent 2.3.344.0  
| | • Upgraded AWS SDK to 3.3.29.13 |
| 4.9.3067 | • Improvements made to instance hibernation  
| | • New version of SSM Agent 2.3.235.0 |
| 4.9.3034 | • Added route 169.254.169.253/32 for DNS server  
| | • New version of SSM Agent 2.3.193.0 |
| 4.9.2986 | • Added signing for all EC2Config related binaries  
| | • New version of SSM Agent 2.3.136.0 |
| 4.9.2953 | New version of SSM Agent (2.3.117.0) |
| 4.9.2926 | New version of SSM Agent (2.3.68.0) |
| 4.9.2905 | • New version of SSM Agent (2.3.50.0)  
| | • Added route 169.254.169.123/32 to AMZN time service  
<p>| | • Added route 169.254.169.249/32 to GRID license service |</p>
<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9.2854</td>
<td>New version of SSM Agent (2.3.13.0)</td>
</tr>
<tr>
<td>4.9.2831</td>
<td>New version of SSM Agent (2.2.916.0)</td>
</tr>
<tr>
<td>4.9.2818</td>
<td>New version of SSM Agent (2.2.902.0)</td>
</tr>
<tr>
<td>4.9.2756</td>
<td>New version of SSM Agent (2.2.800.0)</td>
</tr>
<tr>
<td>4.9.2688</td>
<td>New version of SSM Agent (2.2.607.0)</td>
</tr>
<tr>
<td>4.9.2660</td>
<td>New version of SSM Agent (2.2.546.0)</td>
</tr>
<tr>
<td>4.9.2644</td>
<td>New version of SSM Agent (2.2.493.0)</td>
</tr>
<tr>
<td>4.9.2586</td>
<td>New version of SSM Agent (2.2.392.0)</td>
</tr>
<tr>
<td>4.9.2565</td>
<td>New version of SSM Agent (2.2.355.0)</td>
</tr>
<tr>
<td></td>
<td>• Fixed an issue on M5 and C5 instances (unable to find PV drivers)</td>
</tr>
<tr>
<td></td>
<td>• Add console logging for instance type, newest PV drivers, and NVMe drivers</td>
</tr>
<tr>
<td>4.9.2549</td>
<td>New version of SSM Agent (2.2.325.0)</td>
</tr>
<tr>
<td>4.9.2461</td>
<td>New version of SSM Agent (2.2.257.0)</td>
</tr>
<tr>
<td>4.9.2439</td>
<td>New version of SSM Agent (2.2.191.0)</td>
</tr>
<tr>
<td>4.9.2400</td>
<td>New version of SSM Agent (2.2.160.0)</td>
</tr>
<tr>
<td>4.9.2327</td>
<td>New version of SSM Agent (2.2.120.0)</td>
</tr>
<tr>
<td></td>
<td>• Added COM port discovery on Amazon EC2 bare metal instances</td>
</tr>
<tr>
<td></td>
<td>• Added Hyper-V status logging on Amazon EC2 bare metal instances</td>
</tr>
<tr>
<td>4.9.2294</td>
<td>New version of SSM Agent (2.2.103.0)</td>
</tr>
<tr>
<td>4.9.2262</td>
<td>New version of SSM Agent (2.2.93.0)</td>
</tr>
<tr>
<td>4.9.2246</td>
<td>New version of SSM Agent (2.2.82.0)</td>
</tr>
<tr>
<td>4.9.2218</td>
<td>New version of SSM Agent (2.2.64.0)</td>
</tr>
<tr>
<td>4.9.2212</td>
<td>New version of SSM Agent (2.2.58.0)</td>
</tr>
<tr>
<td>4.9.2203</td>
<td>New version of SSM Agent (2.2.45.0)</td>
</tr>
<tr>
<td>4.9.2188</td>
<td>New version of SSM Agent (2.2.30.0)</td>
</tr>
<tr>
<td>4.9.2180</td>
<td>New version of SSM Agent (2.2.24.0)</td>
</tr>
<tr>
<td></td>
<td>• Added the Elastic GPU plugin for GPU instances</td>
</tr>
<tr>
<td>4.9.2143</td>
<td>New version of SSM Agent (2.2.16.0)</td>
</tr>
<tr>
<td>4.9.2140</td>
<td>New version of SSM Agent (2.1.10.0)</td>
</tr>
</tbody>
</table>

- Fixed an issue causing EBS NVMe volumes to be marked as ephemeral
<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9.2130</td>
<td>New version of SSM Agent (2.1.4.0)</td>
</tr>
<tr>
<td>4.9.2106</td>
<td>New version of SSM Agent (2.0.952.0)</td>
</tr>
<tr>
<td>4.9.2061</td>
<td>New version of SSM Agent (2.0.922.0)</td>
</tr>
<tr>
<td>4.9.2047</td>
<td>New version of SSM Agent (2.0.913.0)</td>
</tr>
<tr>
<td>4.9.2031</td>
<td>New version of SSM Agent (2.0.902.0)</td>
</tr>
<tr>
<td>4.9.2016</td>
<td>• New version of SSM Agent (2.0.879.0)</td>
</tr>
<tr>
<td></td>
<td>• Fixed the CloudWatch Logs directory path for Windows Server 2003</td>
</tr>
<tr>
<td>4.9.1981</td>
<td>• New version of SSM Agent (2.0.847.0)</td>
</tr>
<tr>
<td></td>
<td>• Fixed the issue with important.txt being generated in EBS volumes.</td>
</tr>
<tr>
<td>4.9.1964</td>
<td>New version of SSM Agent (2.0.842.0)</td>
</tr>
<tr>
<td>4.9.1951</td>
<td>• New version of SSM Agent (2.0.834.0)</td>
</tr>
<tr>
<td></td>
<td>• Fixed the issue with drive letter not being mapped from Z: for ephemeral drives.</td>
</tr>
<tr>
<td>4.9.1925</td>
<td>• New version of SSM Agent (2.0.822.0)</td>
</tr>
<tr>
<td></td>
<td>• [Bug] This version is not a valid update target from SSM Agent v4.9.1775.</td>
</tr>
<tr>
<td>4.9.1900</td>
<td>New version of SSM Agent (2.0.805.0)</td>
</tr>
<tr>
<td>4.9.1876</td>
<td>• New version of SSM Agent (2.0.796.0)</td>
</tr>
<tr>
<td></td>
<td>• Fixed an issue with output/error redirection for admin userdata execution.</td>
</tr>
<tr>
<td>4.9.1863</td>
<td>• New version of SSM Agent (2.0.790.0)</td>
</tr>
<tr>
<td></td>
<td>• Fixed problems with attaching multiple EBS volumes to an Amazon EC2 instance.</td>
</tr>
<tr>
<td></td>
<td>• Improved CloudWatch to take a configuration path, keeping the backwards compatibility.</td>
</tr>
<tr>
<td>4.9.1791</td>
<td>New version of SSM Agent (2.0.767.0)</td>
</tr>
<tr>
<td>4.9.1775</td>
<td>New version of SSM Agent (2.0.761.0)</td>
</tr>
<tr>
<td>4.9.1752</td>
<td>New version of SSM Agent (2.0.755.0)</td>
</tr>
<tr>
<td>4.9.1711</td>
<td>New version of SSM Agent (2.0.730.0)</td>
</tr>
<tr>
<td>4.8.1676</td>
<td>New version of SSM Agent (2.0.716.0)</td>
</tr>
<tr>
<td>4.7.1631</td>
<td>New version of SSM Agent (2.0.682.0)</td>
</tr>
<tr>
<td>4.6.1579</td>
<td>• New version of SSM Agent (2.0.672.0)</td>
</tr>
<tr>
<td></td>
<td>• Fixed agent update issue with v4.3, v4.4, and v4.5</td>
</tr>
<tr>
<td>4.5.1534</td>
<td>New version of SSM Agent (2.0.645.1)</td>
</tr>
<tr>
<td>Version</td>
<td>Details</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>4.4.1503</td>
<td>New version of SSM Agent (2.0.633.0)</td>
</tr>
<tr>
<td>4.3.1472</td>
<td>New version of SSM Agent (2.0.617.1)</td>
</tr>
<tr>
<td>4.2.1442</td>
<td>New version of SSM Agent (2.0.599.0)</td>
</tr>
<tr>
<td>4.1.1378</td>
<td>New version of SSM Agent (2.0.558.0)</td>
</tr>
<tr>
<td>4.0.1343</td>
<td>Run Command, State Manager, the CloudWatch agent, and domain join support have been moved into another agent called SSM Agent. SSM Agent will be installed as part of the EC2Config upgrade. For more information, see EC2Config and AWS Systems Manager (p. 408).&lt;br&gt;• If you have a proxy set up in EC2Config, you will need to update your proxy settings for SSM Agent before upgrading. If you do not update the proxy settings, you will not be able to use Run Command to manage your instances. To avoid this, see the following information before updating to the newer version: Installing and Configuring SSM Agent on Windows Instances in the AWS Systems Manager User Guide.&lt;br&gt;• If you previously enabled CloudWatch integration on your instances by using a local configuration file (AWS.EC2.Windows.CloudWatch.json), you will need to configure the file to work with SSM Agent.</td>
</tr>
<tr>
<td>3.19.1153</td>
<td>Re-enabled activation plugin for instances with old KMS configuration.&lt;br&gt;• Change default TRIM behavior to be disabled during disk format operation and added FormatWithTRIM for overriding InitializeDisks plugin with userdata.</td>
</tr>
<tr>
<td>3.18.1118</td>
<td>Fix to reliably add routes to the primary network adapter.&lt;br&gt;• Updates to improve support for AWS services.</td>
</tr>
<tr>
<td>3.17.1032</td>
<td>Fixes duplicate system logs appearing when filters set to same category.&lt;br&gt;• Fixes to prevent from hanging during disk initialization.</td>
</tr>
<tr>
<td>3.16.930</td>
<td>Added support to log &quot;Window is Ready to use&quot; event to Windows Event Log on start.</td>
</tr>
<tr>
<td>3.15.880</td>
<td>Fix to allow uploading Systems Manager Run Command output to S3 bucket names with '.' character.</td>
</tr>
<tr>
<td>Version</td>
<td>Details</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
</tr>
</tbody>
</table>
| 3.14.786 | Added support to override InitializeDisks plugin settings. For example: To speed up SSD disk initialize, you can temporarily disable TRIM by specifying this in userdata:  
  &lt;InitializeDrivesSettings&gt;&lt;SettingsGroup&gt;FormatWithoutTRIM&lt;/SettingsGroup&gt;&lt;/InitializeDrivesSettings |
| 3.13.727 | Systems Manager Run Command - Fixes to process commands reliably after windows reboot. |
| 3.12.649 | • Fix to gracefully handle reboot when running commands/scripts.  
• Fix to reliably cancel running commands.  
• Add support for (optionally) uploading MSI logs to S3 when installing applications via Systems Manager Run Command. |
| 3.11.521 | • Fixes to enable RDP thumbprint generation for Windows Server 2003.  
• Fixes to include timezone and UTC offset in the EC2Config log lines.  
• Systems Manager support to run Run Command commands in parallel.  
• Roll back previous change to bring partitioned disks online. |
| 3.10.442 | • Fix Systems Manager configuration failures when installing MSI applications.  
• Fix to reliably bring storage disks online.  
• Updates to improve support for AWS services. |
| 3.9.359  | • Fix in post Sysprep script to leave the configuration of windows update in a default state.  
• Fix the password generation plugin to improve the reliability in getting GPO password policy settings.  
• Restrict EC2Config/SSM log folder permissions to the local Administrators group.  
• Updates to improve support for AWS services. |
| 3.8.294  | • Fixed an issue with CloudWatch that prevented logs from getting uploaded when not on primary drive.  
• Improved the disk initialization process by adding retry logic.  
• Added improved error handling when the SetPassword plugin occasionally failed during AMI creation.  
• Updates to improve support for AWS services. |
<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7.308</td>
<td>• Improvements to the ec2config-cli utility for config testing and troubleshooting within instance.</td>
</tr>
<tr>
<td></td>
<td>• Avoid adding static routes for KMS and metadata service on an OpenVPN adapter.</td>
</tr>
<tr>
<td></td>
<td>• Fixed an issue where user-data execution was not honoring the &quot;persist&quot; tag.</td>
</tr>
<tr>
<td></td>
<td>• Improved error handling when logging to the EC2 console is not available.</td>
</tr>
<tr>
<td></td>
<td>• Updates to improve support for AWS services.</td>
</tr>
<tr>
<td>3.6.269</td>
<td>• Windows activation reliability fix to first use link local address 169.254.0.250/251 for activating windows via KMS</td>
</tr>
<tr>
<td></td>
<td>• Improved proxy handling for Systems Manager, Windows Activation and Domain Join scenarios</td>
</tr>
<tr>
<td></td>
<td>• Fixed an issue where duplicate lines of user accounts were added to the Sysprep answer file</td>
</tr>
<tr>
<td>3.5.228</td>
<td>• Addressed a scenario where the CloudWatch plugin may consume excessive CPU and memory reading Windows Event Logs</td>
</tr>
<tr>
<td></td>
<td>• Added a link to the CloudWatch configuration documentation in the EC2Config Settings UI</td>
</tr>
<tr>
<td>3.4.212</td>
<td>• Fixes to EC2Config when used in combination with VM-Import.</td>
</tr>
<tr>
<td></td>
<td>• Fixed service naming issue in the WiX installer.</td>
</tr>
<tr>
<td>3.3.174</td>
<td>• Improved exception handling for Systems Manager and domain join failures.</td>
</tr>
<tr>
<td></td>
<td>• Change to support Systems Manager SSM schema versioning.</td>
</tr>
<tr>
<td></td>
<td>• Fixed formatting ephemeral disks on Win2K3.</td>
</tr>
<tr>
<td></td>
<td>• Change to support configuring disk size greater than 2TB. Adam</td>
</tr>
<tr>
<td></td>
<td>• Reduced virtual memory usage by setting GC mode to default. Adam</td>
</tr>
<tr>
<td></td>
<td>• Support for downloading artifacts from UNC path in aws:psModule and aws:application plugin.</td>
</tr>
<tr>
<td></td>
<td>• Improved logging for Windows activation plugin.</td>
</tr>
<tr>
<td>Version</td>
<td>Details</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
</tr>
</tbody>
</table>
| 3.2.97   | • Performance improvements by delay loading Systems Manager SSM assemblies.  
|          | • Improved exception handling for malformed sysprep2008.xml.  
|          | • Command line support for Systems Manager "Apply" configuration.  
|          | • Change to support domain join when there is a pending computer rename.  
|          | • Support for optional parameters in the aws:applications plugin.  
|          | • Support for command array in aws:psModule plugin.  |
| 3.0.54   | • Enable support for Systems Manager.  
|          | • Automatically domain join EC2 Windows instances to an AWS directory via Systems Manager.  
|          | • Configure and upload CloudWatch logs/metrics via Systems Manager.  
|          | • Install PowerShell modules via Systems Manager.  
|          | • Install MSI applications via Systems Manager.  |
| 2.4.233  | • Added scheduled task to recover EC2Config from service startup failures.  
|          | • Improvements to the Console log error messages.  
|          | • Updates to improve support for AWS services.  |
| 2.3.313  | • Fixed an issue with large memory consumption in some cases when the CloudWatch Logs feature is enabled.  
|          | • Fixed an upgrade bug so that ec2config versions lower than 2.1.19 can now upgrade to latest.  
|          | • Updated COM port opening exception to be more friendly and useful in logs.  
|          | • Ec2configServiceSettings UI disabled resizing and fixed the attribution and version display placement in UI.  |
| 2.2.12   | • Handled NullPointerException while querying a registry key for determining Windows Sysprep state which returned null occasionally.  
<p>|          | • Freed up unmanaged resources in finally block.  |
| 2.2.11   | Fixed a issue in CloudWatch plugin for handling empty log lines.  |</p>
<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
</table>
| 2.2.10   | • Removed configuring CloudWatch Logs settings through UI.  
          • Enable users to define CloudWatch Logs settings in %ProgramFiles% \Amazon\Ec2ConfigService\Settings \AWS.EC2.Windows.CloudWatch.json file to allow future enhancements. |
| 2.2.9    | Fixed unhandled exception and added logging. |
| 2.2.8    | • Fixes Windows OS version check in EC2Config Installer to support Windows Server 2003 SP1 and later.  
          • Fixes null value handling when reading registry keys related to updating Sysprep config files. |
| 2.2.7    | • Added support for EC2Config to run during Sysprep execution for Windows 2008 and greater.  
          • Improved exception handling and logging for better diagnostics |
| 2.2.6    | • Reduced the load on the instance and on CloudWatch Logs when uploading log events.  
          • Addressed an upgrade issue where the CloudWatch Logs plug-in did not always stay enabled |
| 2.2.5    | • Added support to upload logs to CloudWatch Log Service.  
          • Fixed a race condition issue in Ec2OutputRDPCert plug-in  
          • Changed EC2Config Service recovery option to Restart from TakeNoAction  
          • Added more exception information when EC2Config Crashes |
| 2.2.4    | • Fixed a typo in PostSysprep.cmd  
          • Fixed the bug which EC2Config does not pin itself onto start menu for OS2012+ |
### Version Details

<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
</table>
| 2.2.3   | - Added option to install EC2Config without service starting immediately upon install. To use, run `Ec2Install.exe start=false` from the command prompt  
- Added parameter in wallpaper plugin to control adding/removing wallpaper. To use, run `Ec2WallpaperInfo.exe set` or `Ec2WallpaperInfo.exe revert` from the command prompt  
- Added checking for RealTimeIsUniversal key, output incorrect settings of the RealTimeIsUniversal registry key to the Console  
- Removed EC2Config dependency on Windows temp folder  
- Removed UserData execution dependency on .Net 3.5 |
| 2.2.2   | - Added check to service stop behavior to check that resources are being released  
- Fixed issue with long execution times when joined to domain |
| 2.2.1   | - Updated Installer to allow upgrades from older versions  
- Fixed Ec2WallpaperInfo bug in .Net4.5 only environment  
- Fixed intermittent driver detection bug  
- Added silent install option. Execute Ec2Install.exe with the `-q` option. eg: `Ec2Install.exe -q` |
| 2.2.0   | - Added support for .Net4 and .Net4.5 only environments  
- Updated Installer |
| 2.1.19  | - Added ephemeral disk labeling support when using Intel network driver (eg. C3 instance Type). For more information, see Enhanced Networking on Windows (p. 700).  
- Added AMI Origin Version and AMI Origin Name support to the console output  
- Made changes to the Console Output for consistent formatting/parsing  
- Updated Help File |
| 2.1.18  | - Added EC2Config WMI Object for Completion notification (-Namespace root\Amazon -Class EC2_ConfigService)  
- Improved Performance of Startup WMI query with large Event Logs; could cause prolonged high CPU during initial execution |
## EC2Config Service

<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
</table>
| 2.1.17  | - Fixed UserData execution issue with Standard Output and Standard Error buffer filling  
- Fixed incorrect RDP thumbprint sometimes appearing in Console Output for >= w2k8 OS  
- Console Output now contains 'RDPCERTIFICATE-SubjectName:' for Windows 2008+, which contains the machine name value  
- Added D:\ to Drive Letter Mapping dropdown  
- Moved Help button to top right and changed look/feel  
- Added Feedback survey link to top right |
| 2.1.16  | - General Tab includes link to EC2Config download page for new Versions  
- Desktop Wallpaper overlay now stored in Users Local Appdata folder instead of My Documents to support MyDoc redirection  
- MSSQLServer name sync'd with system in Post-Sysprep script (2008+)  
- Reordered Application Folder (moved files to Plugin directory and removed duplicate files)  
- Changed System Log Output (Console):  
  - *Moved to a date, name, value format for easier parsing (Please start migrating dependencies to new format)  
  - *Added 'Ec2SetPassword' plugin status  
  - *Added Sysprep Start and End times  
- Fixed issue of Ephemeral Disks not being labeled as 'Temporary Storage' for non-english Operating Systems  
- Fixed EC2Config Uninstall failure after running Sysprep |
| 2.1.15  | - Optimized requests to the Metadata service  
- Metadata now bypass Proxy Settings  
- Ephemeral Disks labeled as 'Temporary Storage' and Important.txt placed on volume when found (Citrix PV drivers only). For more information, see Upgrading PV Drivers on Your Windows Instances (p. 436).  
- Ephemeral Disks assigned drive letters from Z to A (Citrix PV drivers only) - assignment can be overwritten using Drive Letter Mapping plugin with Volume labels 'Temporary Storage X' where x is a number 0-25  
- UserData now executes immediately following 'Windows is Ready' |
<p>| 2.1.14  | Desktop wallpaper fixes |</p>
<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
</table>
| 2.1.13   | • Desktop wallpaper will display hostname by default  
          • Removed dependency on Windows Time service  
          • Route added in cases where multiple IPs are assigned to a single interface |
| 2.1.11   | • Changes made to Ec2Activation Plugin  
          • -Verifies Activation status every 30 days  
          • -If Grace Period has 90 days remaining (out of 180), reattempts activation |
| 2.1.10   | • Desktop wallpaper overlay no longer persists with Sysprep or Shutdown without Sysprep  
          • Userdata option to execute on every service start with `<persist>true</persist>`  
          • Changed location and name of `/DisableWinUpdate.cmd` to `/Scripts/PostSysprep.cmd`  
          • Administrator password set to not expire by default in `/Scripts/PostSysprep.cmd`  
          • Uninstall will remove EC2Config PostSysprep script from `c:\windows\setup\script\CommandComplete.cmd`  
          • Add Route supports custom interface metrics |
| 2.1.9    | UserData Execution no longer limited to 3851 Characters |
| 2.1.7    | • OS Version and language identifier written to console  
          • EC2Config version written to console  
          • PV driver version written to console  
          • Detection of Bug Check and output to the console on next boot when found  
          • Option added to config.xml to persist Sysprep credentials  
          • Add Route Retry logic in cases of ENI being unavailable at start  
          • User Data execution PID written to console  
          • Minimum generated password length retrieved from GPO  
          • Set service start to retry 3 attempts  
          • Added S3_DownloadFile.ps1 and S3_Upload file.ps1 examples to `/Scripts` folder |
<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
</table>
| 2.1.6    | • Version information added to General tab  
                  • Renamed the Bundle tab to Image  
                  • Simplified the process of specifying passwords and moved the password-related UI from the General tab to the Image tab  
                  • Renamed the Disk Settings tab to Storage  
                  • Added a Support tab with common tools for troubleshooting  
                  • Windows Server 2003 sysprep.ini set to extend OS partition by default  
                  • Added the private IP address to the wallpaper  
                  • Private IP address displayed on wallpaper  
                  • Added retry logic for Console output  
                  • Fixed Com port exception for metadata accessibility -- caused EC2Config to terminate before console output is displayed  
                  • Checks for activation status on every boot -- activates as necessary  
                  • Fixed issue of relative paths -- caused when manually executing wallpaper shortcut from startup folder; pointing to Administrator/logs  
                  • Fixed default background color for Windows Server 2003 user (other than Administrator) |
| 2.1.2    | • Console timestamps in UTC (Zulu)  
                  • Removed appearance of hyperlink on Sysprep tab  
                  • Addition of feature to dynamically expand Root Volume on first boot for Windows 2008+  
                  • When Set-Password is enabled, now automatically enables EC2Config to set the password  
                  • EC2Config checks activation status prior to running Sysprep (presents warning if not activated)  
                  • Windows Server 2003 Sysprep.xml now defaults to UTC timezone instead of Pacific  
                  • Randomized Activation Servers  
                  • Renamed Drive Mapping tab to Disk Settings  
                  • Moved Initialize Drives UI items from General to the Disk Settings tab  
                  • Help button now points to HTML help file  
                  • Updated HTML help file with changes  
                  • Updated 'Note' text for Drive Letter Mappings  
                  • Added InstallUpdates.ps1 to /Scripts folder for automating Patches and cleanup prior to Sysprep |
<table>
<thead>
<tr>
<th>Version</th>
<th>Details</th>
</tr>
</thead>
</table>
| 2.1.0   | • Desktop wallpaper displays instance information by default upon first logon (not disconnect/reconnect)  
         | • PowerShell can be executed from the userdata by surrounding the code with <powershell/></powershell> |

Subscribing to EC2Config Service Notifications

Amazon SNS can notify you when new versions of the EC2Config service are released. Use the following procedure to subscribe to these notifications.

To subscribe to EC2Config notifications

2. In the navigation bar, change the region to US East (N. Virginia), if necessary. You must select this region because the SNS notifications that you are subscribing to were created in this region.
3. In the navigation pane, choose Subscriptions.
4. Choose Create subscription.
5. In the Create subscription dialog box, do the following:
   a. For Topic ARN, use the following Amazon Resource Name (ARN):
      
      \[
      \text{arn:aws:sns:us-east-1:801119661308:ec2-windows-ec2config}
      \]
   b. For Protocol, choose Email.
   c. For Endpoint, type an email address that you can use to receive the notifications.
   d. Choose Create subscription.
6. You'll receive a email asking you to confirm your subscription. Open the email and follow the directions to complete your subscription.

Whenever a new version of the EC2Config service is released, we send notifications to subscribers. If you no longer want to receive these notifications, use the following procedure to unsubscribe.

To unsubscribe from EC2Config notifications

1. Open the Amazon SNS console.
2. In the navigation pane, choose Subscriptions.
3. Select the subscription and then choose Actions, Delete subscriptions When prompted for confirmation, choose Delete.

Troubleshooting Issues with the EC2Config Service

The following information can help you troubleshoot issues with the EC2Config service.

Update EC2Config on an Unreachable Instance

Use the following procedure to update the EC2Config service on a Windows Server instance that is inaccessible using Remote Desktop.
To update EC2Config on an Amazon EBS-backed Windows instance that you can't connect to

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Locate the affected instance. Open the context (right-click) menu for the instance, choose Instance State, and then choose Stop.

   **Warning**
   When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.
4. Choose Launch Instance and create a temporary t2.micro instance in the same Availability Zone as the affected instance. Use a different AMI than the one that you used to launch the affected instance.

   **Important**
   If you do not create the instance in the same Availability Zone as the affected instance you will not be able to attach the root volume of the affected instance to the new instance.
5. In the EC2 console, choose Volumes.
6. Locate the root volume of the affected instance. Detach (p. 816) the volume and attach (p. 803) it to the temporary instance that you created earlier. Attach it with the default device name (xvdff).
7. Use Remote Desktop to connect to the temporary instance, and then use the Disk Management utility to make the volume available for use (p. 804).
8. Download the latest version of the EC2Config service. Extract the files from the .zip file to the Temp directory on the drive you attached.
9. On the temporary instance, open the Run dialog box, type regedit, and press Enter.
10. Choose HKEY_LOCAL_MACHINE. From the File menu, choose Load Hive. Choose the drive and then navigate to and open the following file: Windows\System32\config\SOFTWARE. When prompted, specify a key name.
11. Select the key you just loaded and navigate to Microsoft\Windows\CurrentVersion. Choose the RunOnce key. If this key doesn't exist, choose CurrentVersion from the context (right-click) menu, choose New and then choose Key. Name the key RunOnce.
12. From the context (right-click) menu choose the RunOnce key, choose New and then choose String Value. Enter Ec2Install as the name and C:\Temp\Ec2Install.exe /quiet as the data.
13. Choose the HKEY_LOCAL_MACHINE\specified key name\Microsoft\Windows NT \CurrentVersion\Winlogon key. From the context (right-click) menu choose New, and then choose String Value. Enter AutoAdminLogon as the name and 1 as the value data.
14. Choose the HKEY_LOCAL_MACHINE\specified key name\Microsoft\Windows NT \CurrentVersion\Winlogon> key. From the context (right-click) menu choose New, and then choose String Value. Type DefaultUserName as the name and Administrator as the value data.
15. Choose the HKEY_LOCAL_MACHINE\specified key name\Microsoft\Windows NT \CurrentVersion\Winlogon key. From the context (right-click) menu choose New, and then choose String Value. Type DefaultPassword as the name and enter a password in the value data.
16. In the Registry Editor navigation pane, choose the temporary key that you created when you first opened Registry Editor.
17. From the File menu, choose Unload Hive.
18. In Disk Management Utility, choose the drive you attached earlier, open the context (right-click) menu, and choose Offline.
19. In the Amazon EC2 console, detach the affected volume from the temporary instance and reattach it to your instance with the device name /dev/sda1. You must specify this device name to designate the volume as a root volume.
20. Stop and Start Your Instance (p. 383) the instance.
21. After the instance starts, check the system log and verify that you see the message Windows is ready to use.
22. Open Registry Editor and choose HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT \CurrentVersion\Winlogon. Delete the String Value keys you created earlier: AutoAdminLogon, DefaultUserName, and DefaultPassword.

23. Delete or stop the temporary instance you created in this procedure.

Paravirtual Drivers for Windows Instances

Windows AMIs contain a set of drivers to permit access to virtualized hardware. These drivers are used by Amazon EC2 to map instance store and Amazon EBS volumes to their devices. The following table shows key differences between the different drivers.

<table>
<thead>
<tr>
<th></th>
<th>RedHat PV</th>
<th>Citrix PV</th>
<th>AWS PV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instance type</strong></td>
<td>Not supported for all instance types. If you specify an unsupported instance type, the instance is impaired.</td>
<td>Supported for all instance types.</td>
<td>Supported for all instance types.</td>
</tr>
<tr>
<td><strong>Attached volumes</strong></td>
<td>Supports up to 16 attached volumes.</td>
<td>Supports more than 16 attached volumes.</td>
<td>Supports more than 16 attached volumes.</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td>The driver has known issues where the network connection resets under high loads; for example, fast FTP file transfers.</td>
<td>The driver automatically configures jumbo frames on the network adapter when on a compatible instance type. When the instance is in a cluster placement group (p. 710), this offers better network performance between instances in the cluster placement group.</td>
<td></td>
</tr>
</tbody>
</table>

The following list shows which PV drivers you should run on each version of Windows Server on Amazon EC2.

- Windows Server 2019: AWS PV
- Windows Server 2016: AWS PV
- Windows Server 2012 and 2012 R2: AWS PV
The AWS PV drivers are stored in the %ProgramFiles%\Amazon\Xentools directory. This directory also contains public symbols and a command line tool, xenstore_client.exe, that enables you to access entries in XenStore. For example, the following PowerShell command returns the current time from the Hypervisor:

```
PS C:\> [DateTime]::FromFileTimeUTC((gwmi -n root\wmi -cl AWSXenStoreBase).XenTime).ToString("hh:mm:ss")
11:17:00
```

The AWS PV driver components are listed in the Windows registry under HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services. These driver components are as follows: xenbus, xeniface, xennet, xenvbd, and xenvif.

AWS PV drivers also have a Windows service named LiteAgent, which runs in user-mode. It handles tasks such as shutdown and restart events from AWS APIs on Xen generation instances. You can access and manage services by running Services.msc from the command line. When running on Nitro generation instances, the AWS PV drivers are not used and the LiteAgent service will self-stop starting with driver version 8.2.4. Updating to the latest AWS PV driver also updates the LiteAgent and improves reliability on all instance generations.

**Installing the Latest AWS PV Drivers**

Amazon Windows AMIs contain a set of drivers to permit access to virtualized hardware. These drivers are used by Amazon EC2 to map instance store and Amazon EBS volumes to their devices. We recommend that you install the latest drivers to improve stability and performance of your EC2 Windows instances.

**Installation Options**

- You can use AWS Systems Manager to automatically update the PV drivers. For more information, see Walkthrough: Automatically Update PV Drivers on EC2 Windows Instances (Console) in the AWS Systems Manager User Guide.
- You can download the setup package and run the install program manually. For information about downloading and installing the AWS PV drivers, or if you are upgrading a domain controller, see Upgrade Windows Server Instances (AWS PV Upgrade) (p. 436).

**AWS PV Driver Package History**

The following table shows the changes to AWS PV drivers for each driver release.
<table>
<thead>
<tr>
<th>Package version</th>
<th>Details</th>
<th>Release date</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3.2</td>
<td>Enhanced reliability of networking components.</td>
<td>30 July 2019</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Improved performance and robustness of storage component.</td>
<td>12 June 2019</td>
</tr>
<tr>
<td>8.2.7</td>
<td>Improved efficiency to support migrating to latest generation instance types.</td>
<td>20 May 2019</td>
</tr>
<tr>
<td>8.2.6</td>
<td>Improved efficiency of crash dump path.</td>
<td>15 January 2019</td>
</tr>
<tr>
<td>8.2.5</td>
<td>Additional security enhancements. PowerShell installer now available in package.</td>
<td>12 December 2018</td>
</tr>
<tr>
<td>8.2.4</td>
<td>Reliability improvements.</td>
<td>2 October 2018</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Bug fixes and performance improvements. Report EBS volume ID as disk serial number for EBS volumes. This enables cluster scenarios such as S2D.</td>
<td>29 May 2018</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Network and storage performance improvements plus multiple robustness fixes. To verify that this version has been installed, refer to the following Windows registry value: HKLM\Software\Amazon\PVDriver\Version 8.2.1.</td>
<td>8 March 2018</td>
</tr>
<tr>
<td>7.4.6</td>
<td>Stability fixes to make AWS PV drivers more resilient.</td>
<td>26 April 2017</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Added support for Windows Server 2016. Stability fixes for all supported Windows OS versions. *AWS PV driver version 7.4.3’s signature expires on March 29, 2019. We recommend updating to the latest AWS PV driver.</td>
<td>18 Nov 2016</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Stability fixes for support of X1 instance type.</td>
<td>2 Aug 2016</td>
</tr>
<tr>
<td>7.4.1</td>
<td>• Performance improvement in AWS PV Storage driver. • Stability fixes in AWS PV Storage driver: Fixed an issue where the instances were hitting a system crash with bugcheck code 0x0000DEAD. • Stability fixes in AWS PV Network driver. • Added support for Windows Server 2008R2.</td>
<td>12 July 2016</td>
</tr>
<tr>
<td>7.3.2</td>
<td>• Improved logging and diagnostics. • Stability fix in AWS PV Storage driver. In some cases disks may not surface in Windows after reattaching the disk to the instance. • Added support for Windows Server 2012.</td>
<td>24 June 2015</td>
</tr>
<tr>
<td>7.3.1</td>
<td>TRIM update: Fix related to TRIM requests. This fix stabilizes instances and improves instance performance when managing large numbers of TRIM requests.</td>
<td></td>
</tr>
<tr>
<td>Package version</td>
<td>Details</td>
<td>Release date</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>7.3.0</td>
<td>TRIM support: The AWS PV driver now sends TRIM requests to the hypervisor. Ephemeral disks will properly process TRIM requests given the underlying storage supports TRIM (SSD). Note that EBS-based storage does not support TRIM as of March 2015.</td>
<td></td>
</tr>
<tr>
<td>7.2.5</td>
<td>• Stability fix in AWS PV Storage drivers: In some cases the AWS PV driver could dereference invalid memory and cause a system failure.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Stability fix while generating a crash dump: In some cases the AWS PV driver could get stuck in a race condition when writing a crash dump. Before this release, the issue could only be resolved by forcing the driver to stop and restart which lost the memory dump.</td>
<td></td>
</tr>
<tr>
<td>7.2.4</td>
<td>Device ID persistence: This driver fix masks the platform PCI device ID and forces the system to always surface the same device ID, even if the instance is moved. More generally, the fix affects how the hypervisor surfaces virtual devices. The fix also includes modifications to the co-installer for the AWS PV drivers so the system persists mapped virtual devices.</td>
<td></td>
</tr>
<tr>
<td>7.2.2</td>
<td>• Load the AWS PV drivers in Directory Services Restore Mode (DSRM) mode: Directory Services Restore Mode is a safe mode boot option for Windows Server domain controllers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Persist device ID when virtual network adapter device is reattached: This fix forces the system to check the MAC address mapping and persist the device ID. This fix ensures that adapters retain their static settings if the adapters are reattached.</td>
<td></td>
</tr>
<tr>
<td>7.2.1</td>
<td>• Run in safe mode: Fixed an issue where the driver would not load in safe mode. Previously the AWS PV Drivers would only instantiate in normal running systems.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Add disks to Microsoft Windows Storage Pools: Previously we synthesized page 83 queries. The fix disabled page 83 support. Note this does not affect storage pools that are used in a cluster environment because PV disks are not valid cluster disks.</td>
<td></td>
</tr>
<tr>
<td>7.2.0</td>
<td>Base: The AWS PV base version.</td>
<td></td>
</tr>
</tbody>
</table>

**Citrix PV Drivers**

The Citrix PV drivers are stored in the %ProgramFiles%\Citrix\XenTools (32-bit instances) or %ProgramFiles(x86)%\Citrix\XenTools (64-bit instances) directory.

The Citrix PV driver components are listed in the Windows registry under HKEY_LOCAL_MACHINE \SYSTEM\CurrentControlSet\services. These driver components are as follows: xenevtchn, xeniface, xennet, Xennet6, xensvc, xenvbd, and xenvif.

Citrix also has a driver component named XenGuestAgent, which runs as a Windows service. It handles tasks such as shutdown and restart events from the API. You can access and manage services by running Services.msc from the command line.

If you are encountering networking errors while performing certain workloads, you may need to disable the TCP offloading feature for the Citrix PV driver. For more information, see TCP Offloading (p. 445).
RedHat PV Drivers

RedHat drivers are supported for legacy instances, but are not recommended on newer instances with more than 12GB of RAM due to driver limitations. Instances with more than 12GB of RAM running RedHat drivers can fail to boot and become inaccessible. We recommend upgrading RedHat drivers to Citrix PV drivers, and then upgrade Citrix PV drivers to AWS PV drivers.

The source files for the RedHat drivers are in the %ProgramFiles%\RedHat (32-bit instances) or %ProgramFiles(x86)%\RedHat (64-bit instances) directory. The two drivers are rhelnet, the RedHat Paravirtualized network driver, and rhelscsi, the RedHat SCSI miniport driver.

Subscribing to Notifications

Amazon SNS can notify you when new versions of EC2 Windows Drivers are released. Use the following procedure to subscribe to these notifications.

To subscribe to EC2 notifications from the console

2. In the navigation bar, change the region to US East (N. Virginia), if necessary. You must select this region because the SNS notifications that you are subscribing to are in this region.
3. In the navigation pane, choose Subscriptions.
4. Choose Create subscription.
5. In the Create subscription dialog box, do the following:
   a. For TopicARN, copy the following Amazon Resource Name (ARN):
      arn:aws:sns:us-east-1:801119661308:ec2-windows-drivers
   b. For Protocol, choose Email.
   c. For Endpoint, type an email address that you can use to receive the notifications.
   d. Choose Create subscription.
6. You’ll receive a confirmation email. Open the email and follow the directions to complete your subscription.

Whenever new EC2 Windows drivers are released, we send notifications to subscribers. If you no longer want to receive these notifications, use the following procedure to unsubscribe.

To unsubscribe from Amazon EC2 Windows driver notification

2. In the navigation pane, choose Subscriptions.
3. Select the checkbox for the subscription and then choose Actions, Delete subscriptions. When prompted for confirmation, choose Delete.

To subscribe to EC2 notifications using the AWS CLI

To subscribe to EC2 notifications with the AWS CLI, use the following command.

```
aws sns subscribe --topic-arn arn:aws:sns:us-east-1:801119661308:ec2-windows-drivers --protocol email --notification-endpoint YourUserName@YourDomainName.ext
```

To subscribe to EC2 notifications using AWS Tools for Windows PowerShell

To subscribe to EC2 notifications with AWS Tools for Windows PowerShell, use the following command.

```bash
aws sns subscribe --topic-arn arn:aws:sns:us-east-1:801119661308:ec2-windows-drivers --protocol email --notification-endpoint YourUserName@YourDomainName.ext
```
Upgrading PV Drivers on Your Windows Instances

To verify which driver your Windows instance uses, open **Network Connections** in Control Panel and view the **Local Area Connection**. Check whether the driver is one of the following:

- AWS PV Network Device
- Citrix PV Ethernet Adapter
- RedHat PV NIC Driver

Alternatively, you can check the output from the `pnputil -e` command.

**Contents**

- Upgrade Windows Server Instances (AWS PV Upgrade) (p. 436)
- Upgrade a Domain Controller (AWS PV Upgrade) (p. 437)
- Upgrade Windows Server 2008 and 2008 R2 Instances (Redhat to Citrix PV Upgrade) (p. 438)
- Upgrade Your Citrix Xen Guest Agent Service (p. 440)

Upgrade Windows Server Instances (AWS PV Upgrade)

Use the following procedure to perform an in-place upgrade of AWS PV drivers, or to upgrade from Citrix PV drivers to AWS PV drivers on Windows Server 2008 R2, Windows Server 2012, Windows Server 2012 R2, Windows Server 2016, or Windows Server 2019. This upgrade is not available for RedHat drivers, or for other versions of Windows Server.

**Important**

If your instance is a domain controller, see Upgrade a Domain Controller (AWS PV Upgrade) (p. 437). The upgrade process for domain controller instances is different than standard editions of Windows.

**To upgrade AWS PV drivers**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**.
3. Choose the instance that requires the driver upgrade, open the context (right-click) menu, choose **Instance State**, and then choose **Stop**.

**Warning**

When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.

4. After the instance is stopped, create a backup. Open the context (right-click) menu for the instance, choose **Image**, and then choose **Create Image**.
5. From the context (right-click) menu for the instance, choose **Instance State**, and then choose **Start**.
6. Connect to the instance using Remote Desktop and prepare the instance for upgrade. We recommend that you take all non-system disks offline and note any drive letter mappings to the secondary disks in Disk Management before you perform this upgrade. Note that this step is not required if you are performing an in-place update of AWS PV drivers. We also recommend setting non-essential services to **Manual** start-up in the Services console.
7. Download the latest driver package to the instance.
8. Extract the contents of the folder and then run **AWSPVDriverSetup.msi**.
After running the MSI, the instance automatically reboots and then upgrades the driver. The instance will not be available for up to 15 minutes. After the upgrade is complete and the instance passes both health checks in the Amazon EC2 console, you can verify that the new driver was installed by connecting to the instance using Remote Desktop and then running the following PowerShell command:

```
Get-ItemProperty HKLM:\SOFTWARE\Amazon\PVDriver
```

Verify that the driver version is the same as the latest version listed in the Driver Version History table. For more information, see AWS PV Driver Package History (p. 432) Open Disk Management to review any offline secondary volumes and bring them online corresponding to the drive letters noted in Step 6.

If you previously disabled TCP Offloading (p. 445) using Netsh for Citrix PV drivers we recommend that you re-enable this feature after upgrading to AWS PV drivers. TCP Offloading issues with Citrix drivers are not present in the AWS PV drivers. As a result, TCP Offloading provides better performance with AWS PV drivers.

If you previously applied a static IP address or DNS configuration to the network interface, you must reapply the static IP address or DNS configuration after upgrading AWS PV drivers.

**Upgrade a Domain Controller (AWS PV Upgrade)**

Use the following procedure on a domain controller to perform either an in-place upgrade of AWS PV drivers, or to upgrade from Citrix PV drivers to AWS PV drivers.

**To upgrade a domain controller**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Choose the instance that requires the driver upgrade, open the context (right-click) menu, choose **Instance State**, and then choose **Stop**.

   **Warning**
   
   When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.

4. After the instance is stopped, create a backup. Open the context (right-click) menu for the instance, choose **Image**, and then choose **Create Image**.
5. From the context (right-click) menu for the instance, choose **Instance State**, and then choose **Start**.
6. Run the following command to configure Windows to boot into Directory Services Restore Mode (DSRM):

   
   `bcdedit /set {default} safeboot dsrepair`

   **Warning**
   
   Before running this command, confirm that you know the DSRM password. You'll need this information so that you can log in to your instance after the upgrade is complete and the instance automatically reboots.

The system must boot into DSRM because the upgrade utility removes Citrix PV storage drivers so it can install AWS PV drivers. Therefore we recommend noting any drive letter and folder mappings to the secondary disks in Disk Management. When Citrix PV storage drivers are not present, secondary drives will not be detected. Domain controllers that use an NTDS folder on secondary drives will not boot because the secondary disk will not be detected.

**Warning**

After you run this command do not manually reboot the system. The system will be unreachable because Citrix PV drivers do not support DSRM.
7. Run the following command to add **DisableDCCheck** to the registry:

```
reg add HKLM\SOFTWARE\Wow6432Node\Amazon\AWSPVDriverSetup /v DisableDCCheck /t REG_SZ /d true
```

8. **Download** the latest driver package to the instance.

9. Extract the contents of the folder and then run `AWSPVDriverSetup.msi`.

After running the MSI, the instance automatically reboots and then upgrades the driver. The instance will not be available for up to 15 minutes.

10. After the upgrade is complete and the instance passes both health checks in the Amazon EC2 console, connect to the instance using Remote Desktop. Open Disk Management to review any offline secondary volumes and bring them online corresponding to the drive letters and folder mappings noted in Step 6.

    **Important**
    
    You must connect to the instance by specifying the user name in the following format `hostname\administrator`. For example, `Win2k12TestBox\administrator`.

11. Run the following command to remove the DSRM boot configuration:

    ```
bcedit /deletevalue safeboot
    ```

12. Reboot the instance.

13. To complete the upgrade process, verify that the new driver was installed. In Device Manager, under **Storage Controllers**, locate **AWS PV Storage Host Adapter**. Verify that the driver version is the same as the latest version listed in the Driver Version History table. For more information, see **AWS PV Driver Package History** (p. 432).

14. Run the following command to delete **DisableDCCheck** from the registry:

    ```
    reg delete HKLM\SOFTWARE\Wow6432Node\Amazon\AWSPVDriverSetup /v DisableDCCheck
    ```

**Note**

If you previously disabled **TCP Offloading** (p. 445) using Netsh for Citrix PV drivers we recommend that you re-enable this feature after upgrading to AWS PV Drivers. TCP Offloading issues with Citrix drivers are not present in the AWS PV drivers. As a result, TCP Offloading provides better performance with AWS PV drivers.

**Upgrade Windows Server 2008 and 2008 R2 Instances (Redhat to Citrix PV Upgrade)**

Before you start upgrading your RedHat drivers to Citrix PV drivers, make sure you do the following:

- Install the latest version of the EC2Config service. For more information, see **Installing the Latest Version of EC2Config** (p. 406).
- Verify that you have Windows PowerShell 2.0 installed. To verify the version that you have installed, run the following command in a PowerShell window:

  ```
  PS C:\> $PSVersionTable.PSVersion
  ```

  If you need to install version 2.0, see **Windows Management Framework (Windows PowerShell 2.0, WinRM 2.0, and BITS 4.0)** from Microsoft Support.

- Back up your important information on the instance, or create an AMI from the instance. For more information about creating an AMI, see **Creating a Custom Windows AMI** (p. 64). If you create an AMI, make sure that you do the following:
• Write down your password.
• Do not run the Sysprep tool manually or using the EC2Config service.
• Set your Ethernet adapter to obtain an IP address automatically using DHCP. For more information, see Configure TCP/IP Settings in the Microsoft TechNet Library.

To upgrade Redhat drivers
1. Connect to your instance and log in as the local administrator. For more information about connecting to your instance, see Connecting to Your Windows Instance (p. 379).
2. In your instance, download the Citrix PV upgrade package.
3. Extract the contents of the upgrade package to a location of your choice.
4. Double-click the Upgrade.bat file. If you get a security warning, choose Run.
5. In the Upgrade Drivers dialog box, review the information and choose Yes if you are ready to start the upgrade.
6. In the Red Hat Paravirtualized Xen Drivers for Windows uninstaller dialog box, choose Yes to remove the RedHat software. Your instance will be rebooted.
   
   **Note**
   If you do not see the uninstaller dialog box, choose Red Hat Paravirtualize in the Windows taskbar.
7. Check that the instance has rebooted and is ready to be used.
   a. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
   b. On the Instances page, right-click your instance and select Get System Log.
   c. The upgrade operations should have restarted the server 3 or 4 times. You can see this in the log file by the number of times Windows is Ready to use is displayed.
8. Connect to your instance and log in as the local administrator.
10. Confirm that the installation is complete. Navigate to the Citrix-WIN_PV folder that you extracted earlier, open the PVUpgrade.log file, and then check for the text INSTALLATION IS COMPLETE.
Upgrade Your Citrix Xen Guest Agent Service

If you are using Citrix PV drivers on Windows Server, you can upgrade the Citrix Xen guest agent service. This Windows service handles tasks such as shutdown and restart events from the API. You can run this upgrade package on any version of Windows Server, as long as the instance is running Citrix PV drivers.

**Important**
For Windows Server 2008 R2 and later, we recommend you upgrade to AWS PV drivers that include the Guest Agent update.

Before you start upgrading your drivers, make sure you back up your important information on the instance, or create an AMI from the instance. For more information about creating an AMI, see *Creating a Custom Windows AMI* (p. 64). If you create an AMI, make sure you do the following:

- Do not enable the Sysprep tool in the EC2Config service.
- Write down your password.
- Set your Ethernet adapter to DHCP.

**To upgrade your Citrix Xen guest agent service**

1. Connect to your instance and log in as the local administrator. For more information about connecting to your instance, see *Connecting to Your Windows Instance* (p. 379).
2. On your instance, download the Citrix upgrade package.
3. Extract the contents of the upgrade package to a location of your choice.
4. Double-click the `Upgrade.bat` file. If you get a security warning, choose **Run**.
5. In the **Upgrade Drivers** dialog box, review the information and choose **Yes** if you are ready to start the upgrade.
6. When the upgrade is complete, the `PVUpgrade.log` file will open and contain the text **UPGRADE IS COMPLETE**.
7. Reboot your instance.
Troubleshooting PV Drivers

The following are solutions to issues that you might encounter with older Amazon EC2 images and PV drivers.

Contents
- Windows Server 2012 R2 loses network and storage connectivity after an instance reboot (p. 441)
- TCP Offloading (p. 445)
- Time Synchronization (p. 445)

Windows Server 2012 R2 loses network and storage connectivity after an instance reboot

Important
This issue occurs only with AMIs made available before September 2014.

Windows Server 2012 R2 Amazon Machine Images (AMIs) made available before September 10, 2014 can lose network and storage connectivity after an instance reboot. The error in the AWS Management Console system log states: “Difficulty detecting PV driver details for Console Output.” The connectivity loss is caused by the Plug and Play Cleanup feature. This features scans for and disables inactive system devices every 30 days. The feature incorrectly identifies the EC2 network device as inactive and removes it from the system. When this happens, the instance loses network connectivity after a reboot.

For systems that you suspect could be affected by this issue, you can download and run an in-place driver upgrade. If you are unable to perform the in-place driver upgrade, you can run a helper script. The script determines if your instance is affected. If it is affected, and the Amazon EC2 network device has not been removed, the script disables the Plug and Play Cleanup scan. If the network device was removed, the script repairs the device, disables the Plug and Play Cleanup scan, and enables your instance to reboot with network connectivity enabled.

Contents
- Choose How You Want to Fix This Problem (p. 441)
- Method 1 - Enhanced Networking (p. 442)
- Method 2 - Registry configuration (p. 442)
- Run the Remediation Script (p. 444)

Choose How You Want to Fix This Problem

There are two methods for restoring network and storage connectivity to an instance affected by this issue. Choose one of the following methods:

<table>
<thead>
<tr>
<th>Method</th>
<th>Prerequisites</th>
<th>Procedure Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1 - Enhanced networking</td>
<td>Enhanced networking is only available in a virtual private cloud (VPC) which requires a C3 instance type. If the server does not currently use the C3 instance type, then you must temporarily change it.</td>
<td>You change the server instance type to a C3 instance. Enhanced networking then enables you to connect to the affected instance and fix the problem. After you fix the problem, you change the instance back to the original instance type. This method is typically faster than Method 2 and less likely to</td>
</tr>
<tr>
<td>Method</td>
<td>Prerequisites</td>
<td>Procedure Overview</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Method 1 - Enhanced Networking</td>
<td></td>
<td>result in user error. You will incur additional charges as long as the C3 instance is running.</td>
</tr>
<tr>
<td>Method 2 - Registry configuration</td>
<td>Ability to create or access a second server. Ability to change Registry settings.</td>
<td>You detach the root volume from the affected instance, attach it to a different instance, connect, and make changes in the Registry. You will incur additional charges as long as the additional server is running. This method is slower than Method 1, but this method has worked in situations where Method 1 failed to resolve the problem.</td>
</tr>
</tbody>
</table>

**Method 1 - Enhanced Networking**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Locate the affected instance. Open the context (right-click) menu for the instance, choose **Instance State**, and then choose **Stop**.
   
   **Warning**
   When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.

4. After the instance is stopped create a backup. Open the context (right-click) menu for the instance, choose **Image**, and then choose **Create Image**.
5. Change the instance type to any C3 instance type.
6. Start the instance.
7. Connect to the instance using Remote Desktop and then download the AWS PV Drivers Upgrade package to the instance.
8. Extract the contents of the folder and run AWSPVDriverSetup.msi.
   
   After running the MSI, the instance automatically reboots and then upgrades the drivers. The instance will not be available for up to 15 minutes.

9. After the upgrade is complete and the instance passes both health checks in the Amazon EC2 console, connect to the instance using Remote Desktop and verify that the new drivers were installed. In Device Manager, under **Storage Controllers**, locate **AWS PV Storage Host Adapter**. Verify that the driver version is the same as the latest version listed in the Driver Version History table. For more information, see AWS PV Driver Package History (p. 432).
10. Stop the instance and change the instance back to its original instance type.
11. Start the instance and resume normal use.

**Method 2 - Registry configuration**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Locate the affected instance. Open the context (right-click) menu for the instance, choose **Instance State**, and then choose **Stop**.
Warning
When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.


Important
If you do not create the instance in the same Availability Zone as the affected instance you will not be able to attach the root volume of the affected instance to the new instance.

5. In the navigation pane, choose Volumes.

6. Locate the root volume of the affected instance. Detach (p. 816) the volume and attach (p. 803) it to the temporary instance you created earlier. Attach it with the default device name (xvdf).

7. Use Remote Desktop to connect to the temporary instance, and then use the Disk Management utility to make the volume available for use (p. 804).

8. On the temporary instance, open the Run dialog box, type regedit, and press Enter.

9. In the Registry Editor navigation pane, choose HKEY_Local_Machine, and then from the File menu choose Load Hive.

10. In the Load Hive dialog box, navigate to Affected Volume\Windows\System32\config\System and type a temporary name in the Key Name dialog box. For example, enter OldSys.

11. In the navigation pane of the Registry Editor, locate the following keys:

HKEY_LOCAL_MACHINE\your_temporary_key_name\ControlSet001\Control\Class\4d36e97d-e325-11ce-bfc1-08002be10318

HKEY_LOCAL_MACHINE\your_temporary_key_name\ControlSet001\Control\Class\4d36e96a-e325-11ce-bfc1-08002be10318

12. For each key, double-click UpperFilters, enter a value of XENFILT, and then choose OK.

13. Locate the following key:

HKEY_LOCAL_MACHINE\your_temporary_key_name\ControlSet001\Services\XENBUS\Parameters

14. Create a new string (REG_SZ) with the name ActiveDevice and the following value:

PCI\VEN_5B53&DEV_0001&SUBSYS_00015B53&REV_01

15. Locate the following key:

HKEY_LOCAL_MACHINE\your_temporary_key_name\ControlSet001\Services\XENBUS
16. Change the Count from 0 to 1.
17. Locate and delete the following keys:

\HKEY_LOCAL_MACHINE\your_temporary_key_name\ControlSet001\Services\xenvbd
\StartOverride

\HKEY_LOCAL_MACHINE\your_temporary_key_name\ControlSet001\Services\xenfilt
\StartOverride

18. In the Registry Editor navigation pane, choose the temporary key that you created when you first opened the Registry Editor.
19. From the File menu, choose Unload Hive.
20. In the Disk Management Utility, choose the drive you attached earlier, open the context (right-click) menu, and choose Offline.
21. In the Amazon EC2 console, detach the affected volume from the temporary instance and reattach it to your Windows Server 2012 R2 instance with the device name /dev/sda1. You must specify this device name to designate the volume as a root volume.
22. Start the instance.
23. Connect to the instance using Remote Desktop and then download the AWS PV Drivers Upgrade package to the instance.
24. Extract the contents of the folder and run AWSPVDriverSetup.msi.

After running the MSI, the instance automatically reboots and then upgrades the drivers. The instance will not be available for up to 15 minutes.
25. After the upgrade is complete and the instance passes both health checks in the Amazon EC2 console, connect to the instance using Remote Desktop and verify that the new drivers were installed. In Device Manager, under Storage Controllers, locate AWS PV Storage Host Adapter. Verify that the driver version is the same as the latest version listed in the Driver Version History table. For more information, see AWS PV Driver Package History (p. 432).
26. Delete or stop the temporary instance you created in this procedure.

Run the Remediation Script

If you are unable to perform an in-place driver upgrade or migrate to a newer instance you can run the remediation script to fix the problems caused by the Plug and Play Cleanup task.

To run the remediation script

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Choose the instance for which you want to run the remediation script. Open the context (right-click) menu for the instance, choose Instance State, and then choose Stop.

   Warning
   When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.
4. After the instance is stopped create a backup. Open the context (right-click) menu for the instance, choose Image, and then choose Create Image.
5. Open the context (right-click) menu for the instance, choose Instance State, and then choose Start.
6. Connect to the instance by using Remote Desktop and then download the RemediateDriverIssue.zip folder to the instance.
7. Extract the contents of the folder.
8. Run the remediation script according to the instructions in the Readme.txt file. The file is located in the folder where you extracted RemediateDriverIssue.zip.
TCP Offloading

**Important**
This issue does not apply to instances running AWS PV or Intel network drivers.

By default, TCP offloading is enabled for the Citrix PV drivers in Windows AMIs. If you encounter transport-level errors or packet transmission errors (as visible on the Windows Performance Monitor)—for example, when you’re running certain SQL workloads—you may need to disable this feature.

**Warning**
Disabling TCP offloading may reduce the network performance of your instance.

To disable TCP offloading for Windows Server 2012 and 2008

1. Connect to your instance and log in as the local administrator.
2. If you're using Windows Server 2012, press Ctrl+Esc to access the Start screen, and then choose Control Panel. If you're using Windows Server 2008, choose Start and select Control Panel.
3. Choose Network and Internet, then Network and Sharing Center.
5. Right-click Citrix PV Ethernet Adapter #0 and select Properties.

6. In the Local Area Connection Properties dialog box, choose Configure to open the Citrix PV Ethernet Adapter #0 Properties dialog box.
7. On the Advanced tab, disable each of the properties, except for Correct TCP/UDP Checksum Value. To disable a property, select it from Property and choose Disabled from Value.
8. Choose OK.
9. Run the following commands from a Command Prompt window.

```bash
netsh int ip set global taskoffload=disabled
netsh int tcp set global chimney=disabled
netsh int tcp set global rss=disabled
netsh int tcp set global netdma=disabled
```
10. Reboot the instance.

Time Synchronization

Prior to the release of the 2013.02.13 Windows AMI, the Citrix Xen guest agent could set the system time incorrectly. This can cause your DHCP lease to expire. If you have issues connecting to your instance, you might need to update the agent.
To determine whether you have the updated Citrix Xen guest agent, check whether the `C:\Program Files\Citrix\XenGuestAgent.exe` file is from March 2013. If the date on this file is earlier than that, update the Citrix Xen guest agent service. For more information, see Upgrade Your Citrix Xen Guest Agent Service (p. 440).

AWS NVMe Drivers for Windows Instances

EBS volumes and instance store volumes are exposed as NVMe block devices on Nitro-based instances (p. 122). You must have the AWS NVMe driver installed in order to use an NVMe block device. The latest AWS Windows AMIs for Windows Server 2008 R2 and later contain the required AWS NVMe driver.

For more information about EBS and NVMe, see Amazon EBS and NVMe on Windows Instances (p. 875). For more information about SSD instance store and NVMe, see SSD Instance Store Volumes (p. 924).

Installing or Upgrading AWS NVMe Drivers

If you are not using the latest AWS Windows AMIs provided by Amazon, use the following procedure to install the current AWS NVMe driver.

To download and install the latest AWS NVMe driver

1. Connect to your instance and log in as the local administrator.
2. Download the latest driver package to the instance.
3. Extract the zip archive.
4. Install the driver by running the `install.ps1` PowerShell script.
5. If the installer does not reboot your instance for you, restart the instance.

AWS NVMe Driver Version History

The following table describes the released versions of the AWS NVMe driver.

<table>
<thead>
<tr>
<th>Driver version</th>
<th>Details</th>
<th>Release date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3.2</td>
<td>Fixed issue with modifying EBS volumes actively processing IO, which may result in data corruption. Customers who do not modify online EBS volumes (for example, resizing or changing type) are not impacted.</td>
<td>10 September 2019</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Reliability Improvements</td>
<td>21 May 2019</td>
</tr>
<tr>
<td>1.3.0</td>
<td>Device optimization improvements</td>
<td>31 August 2018</td>
</tr>
<tr>
<td>1.2.0</td>
<td>Performance and reliability improvements for AWS NVMe devices on all supported instances, including bare metal instances</td>
<td>13 June 2018</td>
</tr>
<tr>
<td>1.0.0</td>
<td>AWS NVMe driver for supported instance types running Windows Server</td>
<td>12 February 2018</td>
</tr>
</tbody>
</table>

Subscribing to Notifications

Amazon SNS can notify you when new versions of EC2 Windows Drivers are released. Use the following procedure to subscribe to these notifications.
To subscribe to EC2 notifications from the console

2. In the navigation bar, change the region to US East (N. Virginia), if necessary. You must select this region because the SNS notifications that you are subscribing to are in this region.
3. In the navigation pane, choose Subscriptions.
4. Choose Create subscription.
5. In the Create subscription dialog box, do the following:
   a. For TopicARN, copy the following Amazon Resource Name (ARN):
      arn:aws:sns:us-east-1:801119661308:ec2-windows-drivers
   b. For Protocol, choose Email.
   c. For Endpoint, type an email address that you can use to receive the notifications.
   d. Choose Create subscription.
6. You’ll receive a confirmation email. Open the email and follow the directions to complete your subscription.

Whenever new EC2 Windows drivers are released, we send notifications to subscribers. If you no longer want to receive these notifications, use the following procedure to unsubscribe.

To unsubscribe from Amazon EC2 Windows driver notification

2. In the navigation pane, choose Subscriptions.
3. Select the checkbox for the subscription and then choose Actions, Delete subscriptions. When prompted for confirmation, choose Delete.

To subscribe to EC2 notifications using the AWS CLI

To subscribe to EC2 notifications with the AWS CLI, use the following command.

```
aws sns subscribe --topic-arn arn:aws:sns:us-east-1:801119661308:ec2-windows-drivers --protocol email --notification-endpoint YourUserName@YourDomainName.ext
```

To subscribe to EC2 notifications using AWS Tools for Windows PowerShell

To subscribe to EC2 notifications with AWS Tools for Windows PowerShell, use the following command.

```
Connect-SNSNotification -TopicArn 'arn:aws:sns:us-east-1:801119661308:ec2-windows-drivers' -Protocol email -Region us-east-1 -Endpoint 'YourUserName@YourDomainName.ext'
```

Optimizing CPU Options

Amazon EC2 instances support multithreading, which enables multiple threads to run concurrently on a single CPU core. Each thread is represented as a virtual CPU (vCPU) on the instance. An instance has a default number of CPU cores, which varies according to instance type. For example, an m5.xlarge instance type has two CPU cores and two threads per core by default—four vCPUs in total.

**Note**

Each vCPU is a thread of a CPU core, except for T2 instances.

In most cases, there is an Amazon EC2 instance type that has a combination of memory and number of vCPUs to suit your workloads. However, you can specify the following CPU options to optimize your instance for specific workloads or business needs:
• **Number of CPU cores**: You can customize the number of CPU cores for the instance. You might do this to potentially optimize the licensing costs of your software with an instance that has sufficient amounts of RAM for memory-intensive workloads but fewer CPU cores.

• **Threads per core**: You can disable multithreading by specifying a single thread per CPU core. You might do this for certain workloads, such as high performance computing (HPC) workloads.

You can specify these CPU options during instance launch. There is no additional or reduced charge for specifying CPU options. You’re charged the same as instances that are launched with default CPU options.

**Contents**

- Rules for Specifying CPU Options (p. 448)
- CPU Cores and Threads Per CPU Core Per Instance Type (p. 448)
- Specifying CPU Options for Your Instance (p. 458)
- Viewing the CPU Options for Your Instance (p. 459)

**Rules for Specifying CPU Options**

To specify the CPU options for your instance, be aware of the following rules:

• CPU options can only be specified during instance launch and cannot be modified after launch.

• When you launch an instance, you must specify both the number of CPU cores and threads per core in the request. For example requests, see Specifying CPU Options for Your Instance (p. 458).

• The number of vCPUs for the instance is the number of CPU cores multiplied by the threads per core. To specify a custom number of vCPUs, you must specify a valid number of CPU cores and threads per core for the instance type. You cannot exceed the default number of vCPUs for the instance. For more information, see CPU Cores and Threads Per CPU Core Per Instance Type (p. 448).

• To disable multithreading, specify one thread per core.

• When you change the instance type (p. 188) of an existing instance, the CPU options automatically change to the default CPU options for the new instance type.

• The specified CPU options persist after you stop, start, or reboot an instance.

**CPU Cores and Threads Per CPU Core Per Instance Type**

The following tables list the instance types that support specifying CPU options. For each type, the table shows the default and supported number of CPU cores and threads per core.

**Accelerated Computing Instances**

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>f1.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>f1.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>f1.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Default CPU cores</td>
<td>Default threads per core</td>
<td>Valid number of CPU cores</td>
<td>Valid number of threads per core</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>g3.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>g3.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>g3.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>g3s.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>g4dn.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>g4dn.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>g4dn.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>g4dn.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>g4dn.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>g4dn.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>p2.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>p2.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>p2.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>p3.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>p3.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
</tbody>
</table>
### Optimizing CPU Options

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>p3.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>p3dn.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 34, 36, 38, 40, 42, 44, 46, 48</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

### Compute Optimized Instances

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>c4.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>c4.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>c4.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>c4.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>c4.8xlarge</td>
<td>36</td>
<td>18</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.9xlarge</td>
<td>36</td>
<td>18</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.18xlarge</td>
<td>72</td>
<td>36</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36</td>
<td>1, 2</td>
</tr>
</tbody>
</table>
### Amazon Elastic Compute Cloud
#### User Guide for Windows Instances
#### Optimizing CPU Options

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>c5d.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5d.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5d.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5d.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5d.9xlarge</td>
<td>36</td>
<td>18</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5d.18xlarge</td>
<td>72</td>
<td>36</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5n.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5n.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5n.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5n.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5n.9xlarge</td>
<td>36</td>
<td>18</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18</td>
<td>1, 2</td>
</tr>
<tr>
<td>c5n.18xlarge</td>
<td>72</td>
<td>36</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

**General Purpose Instances**

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>m5.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
</tbody>
</table>
### Optimizing CPU Options

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>m5.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 36, 38, 40, 42, 44, 46, 48</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>6, 12, 18, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5a.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>12, 18, 24, 36, 48</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5ad.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5ad.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5ad.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5ad.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5ad.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>6, 12, 18, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5ad.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>12, 18, 24, 36, 48</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5d.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5d.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5d.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5d.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5d.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16</td>
<td>1, 2</td>
</tr>
</tbody>
</table>
### Optimizing CPU Options

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>m5d.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5d.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>m5d.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3.nano</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3.micro</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3.small</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3.medium</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3a.nano</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3a.micro</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3a.small</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3a.medium</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3a.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3a.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>t3a.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

### Memory Optimized Instances

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>r4.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r4.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>r4.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Default CPU cores</td>
<td>Default threads per core</td>
<td>Valid number of CPU cores</td>
<td>Valid number of threads per core</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>r4.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>r4.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>r4.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5.1large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5.2xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5.4xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5.8xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5.12xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5.16xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5.24xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5a.1large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5a.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5a.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5a.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>2, 4, 6, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5a.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5a.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>6, 12, 18, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Default CPU cores</td>
<td>Default threads per core</td>
<td>Valid number of CPU cores</td>
<td>Valid number of threads per core</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>r5a.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5a.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>12, 18, 24, 36, 48</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5ad.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>24</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5ad.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>24</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5ad.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>24</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5ad.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>12, 18, 24, 36, 48</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5ad.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>12, 18, 24</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5ad.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>12, 18, 24, 36, 48</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>r5d.24xlarge</td>
<td>96</td>
<td>48</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>x1.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
</tbody>
</table>
### Optimizing CPU Options

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1.32xlarge</td>
<td>128</td>
<td>64</td>
<td>2</td>
<td>4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64</td>
<td>1, 2</td>
</tr>
<tr>
<td>x1e.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>x1e.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>x1e.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>x1e.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>x1e.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>x1e.32xlarge</td>
<td>128</td>
<td>64</td>
<td>2</td>
<td>4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64</td>
<td>1, 2</td>
</tr>
<tr>
<td>z1d.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>z1d.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>z1d.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>z1d.3xlarge</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>2, 4, 6</td>
<td>1, 2</td>
</tr>
<tr>
<td>z1d.6xlarge</td>
<td>24</td>
<td>12</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12</td>
<td>1, 2</td>
</tr>
<tr>
<td>z1d.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

### Storage Optimized Instances

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Default vCPUs</th>
<th>Default CPU cores</th>
<th>Default threads per core</th>
<th>Valid number of CPU cores</th>
<th>Valid number of threads per core</th>
</tr>
</thead>
<tbody>
<tr>
<td>d2.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>d2.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>d2.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>Instance type</td>
<td>Default vCPUs</td>
<td>Default CPU cores</td>
<td>Default threads per core</td>
<td>Valid number of CPU cores</td>
<td>Valid number of threads per core</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>d2.8xlarge</td>
<td>36</td>
<td>18</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18</td>
<td>1, 2</td>
</tr>
<tr>
<td>h1.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>h1.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>h1.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>h1.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>1, 2, 3, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3.4xlarge</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3.8xlarge</td>
<td>32</td>
<td>16</td>
<td>2</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3.16xlarge</td>
<td>64</td>
<td>32</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3en.large</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3en.xlarge</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3en.2xlarge</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>2, 4</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3en.3xlarge</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>2, 4, 6</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3en.6xlarge</td>
<td>24</td>
<td>12</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12</td>
<td>1, 2</td>
</tr>
<tr>
<td>i3en.12xlarge</td>
<td>48</td>
<td>24</td>
<td>2</td>
<td>2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24</td>
<td>1, 2</td>
</tr>
</tbody>
</table>
### Specifying CPU Options for Your Instance

You can specify CPU options during instance launch. The following examples are for an `r4.4xlarge` instance type, which has the following default values (p. 453):

- Default CPU cores: 8
- Default threads per core: 2
- Default vCPUs: 16 (8 * 2)
- Valid number of CPU cores: 1, 2, 3, 4, 5, 6, 7, 8
- Valid number of threads per core: 1, 2

#### Disabling Multithreading

To disable multithreading, specify one thread per core.

**To disable multithreading during instance launch (console)**

1. Follow the Launching an Instance Using the Launch Instance Wizard (p. 334) procedure.
2. On the Configure Instance Details page, for CPU options, choose Specify CPU options.
3. For Core count, choose the number of required CPU cores. In this example, to specify the default CPU core count for an r4.4xlarge instance, choose 8.
4. To disable multithreading, for Threads per core, choose 1.
5. Continue as prompted by the wizard. When you've finished reviewing your options on the Review Instance Launch page, choose Launch. For more information, see Launching an Instance Using the Launch Instance Wizard (p. 334).

**To disable multithreading during instance launch (AWS CLI)**

Use the `run-instances` AWS CLI command and specify a value of 1 for ThreadsPerCore for the `--cpu-options` parameter. For CoreCount, specify the number of CPU cores. In this example, to specify the default CPU core count for an r4.4xlarge instance, specify a value of 8.

```bash
aws ec2 run-instances --image-id ami-1a2b3c4d --instance-type r4.4xlarge --cpu-options "CoreCount=8,ThreadsPerCore=1" --key-name MyKeyPair
```

#### Specifying a Custom Number of vCPUs

You can customize the number of CPU cores and threads per core for the instance.

**To specify a custom number of vCPUs during instance launch (console)**

The following example launches an r4.4xlarge instance with six vCPUs.
1. Follow the Launching an Instance Using the Launch Instance Wizard (p. 334) procedure.
2. On the Configure Instance Details page, for CPU options, choose Specify CPU options.
3. To get six vCPUs, specify three CPU cores and two threads per core, as follows:
   - For Core count, choose 3.
   - For Threads per core, choose 2.
4. Continue as prompted by the wizard. When you’ve finished reviewing your options on the Review Instance Launch page, choose Launch. For more information, see Launching an Instance Using the Launch Instance Wizard (p. 334).

To specify a custom number of vCPUs during instance launch (AWS CLI)

The following example launches an r4.4xlarge instance with six vCPUs.

Use the run-instances AWS CLI command and specify the number of CPU cores and number of threads in the --cpu-options parameter. You can specify three CPU cores and two threads per core to get six vCPUs.

```bash
aws ec2 run-instances --image-id ami-1a2b3c4d --instance-type r4.4xlarge --cpu-options "CoreCount=3,ThreadsPerCore=2" --key-name MyKeyPair
```

Alternatively, specify six CPU cores and one thread per core (disable multithreading) to get six vCPUs:

```bash
aws ec2 run-instances --image-id ami-1a2b3c4d --instance-type r4.4xlarge --cpu-options "CoreCount=6,ThreadsPerCore=1" --key-name MyKeyPair
```

Viewing the CPU Options for Your Instance

You can view the CPU options for an existing instance in the Amazon EC2 console or by describing the instance using the AWS CLI.

To view the CPU options for an instance (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the left navigation pane, choose Instances, and select the instance.
3. Choose Description and view the Number of vCPUs field.
4. To view the core count and threads per core, choose the Number of vCPUs field value.

To view the CPU options for an instance (AWS CLI)

Use the describe-instances command.

```bash
aws ec2 describe-instances --instance-ids i-123456789abcdef123
```

```json
...
   "Instances": [  
      {  
         "Monitoring": {  
            "State": "disabled"  
         },  
         "PublicDnsName": "ec2-198-51-100-5.eu-central-1.compute.amazonaws.com",  
         "State": {  
            "Code": 16,
```
In the output that's returned, the `CoreCount` field indicates the number of cores for the instance. The `ThreadsPerCore` field indicates the number of threads per core.

Alternatively, connect to your instance and use Task Manager to view the CPU information for your instance.

You can use AWS Config to record, assess, audit, and evaluate configuration changes for instances, including terminated instances. For more information, see Getting Started with AWS Config in the AWS Config Developer Guide.

**Setting the Time for a Windows Instance**

A consistent and accurate time reference is crucial for many server tasks and processes. Most system logs include a time stamp that you can use to determine when problems occur and in what order the events take place. If you use the AWS CLI or an AWS SDK to make requests from your instance, these tools sign requests on your behalf. If your instance's date and time are not set correctly, the date in the signature may not match the date of the request, and AWS rejects the request. We recommend that you use Coordinated Universal Time (UTC) for your Windows instances. However, you can use a different time zone if you want.

**Contents**

- Changing the Time Zone (p. 460)
- Configuring Network Time Protocol (NTP) (p. 461)
- Default Network Time Protocol (NTP) Settings for Amazon Windows AMIs (p. 462)
- Configuring Time Settings for Windows Server 2008 and later (p. 463)
- Related Topics (p. 463)

**Changing the Time Zone**

Windows instances are set to the UTC time zone by default. You can change the time to correspond to your local time zone or a time zone for another part of your network.

**To change the time zone on an instance**

1. From your instance, open a Command Prompt window.
2. Identify the time zone to use on the instance. To get a list of time zones, use the following command: `tzutil /l`. This command returns a list of all available time zones, using the following format:
3. Locate the time zone ID to assign to the instance.
4. Assign the time zone to the instance by using the following command:

```text
    tzutil /s "Pacific Standard Time"
```

The new time zone should take effect immediately.

## Configuring Network Time Protocol (NTP)

Amazon provides the Amazon Time Sync Service, which is accessible from all EC2 instances, and is also used by other AWS services. We recommend that you configure your instance to use the Amazon Time Sync Service. This service uses a fleet of satellite-connected and atomic reference clocks in each AWS Region to deliver accurate current time readings of the Coordinated Universal Time (UTC) global standard. The Amazon Time Sync Service automatically smooths any leap seconds that are added to UTC. This service is available at the `169.254.169.123` IP address for any instance running in a VPC, and your instance does not require internet access to use it. Starting with the August 2018 release, Windows AMIs use the Amazon Time Sync Service by default.

### To verify the NTP configuration

1. From your instance, open a Command Prompt window.
2. Get the current NTP configuration by typing the following command:

```text
    w32tm /query /configuration
```

This command returns the current configuration settings for the Windows instance.
3. (Optional) Get the status of the current configuration by typing the following command:

```text
    w32tm /query /status
```

This command returns information such as the last time the instance synced with the NTP server and the poll interval.

### To change the NTP server to use the Amazon Time Sync Service

1. From the Command Prompt window, run the following command:

```text
    w32tm /config /manualpeerlist:169.254.169.123 /syncfromflags:manual /update
```
2. Verify your new settings by using the following command:

```text
    w32tm /query /configuration
```

In the output that's returned, verify that `NtpServer` displays the `169.254.169.123` IP address.

You can change the instance to use a different set of NTP servers if required. For example, if you have Windows instances that do not have internet access, you can configure them to use an NTP server located within your private network. If your instance is within a domain, you should change the settings...
to use the domain controllers as the time source to avoid time skew. The security group of your instance must be configured to allow outbound UDP traffic on port 123 (NTP).

**To change the NTP servers**

1. From the Command Prompt window, run the following command:

   ```
   w32tm /config /manualpeerlist:comma-delimited list of NTP servers /syncfromflags:manual /update
   ```

   Where *comma-delimited list of NTP servers* is the list of NTP servers for the instance to use.

2. Verify your new settings by using the following command:

   ```
   w32tm /query /configuration
   ```

**Default Network Time Protocol (NTP) Settings for Amazon Windows AMIs**

Amazon Machine Images (AMIs) generally adhere to the out-of-the-box defaults except in cases where changes are required to function on EC2 infrastructure. The following settings have been determined to work well in a virtual environment, as well as to keep any clock drift to within one second of accuracy:

- **Update Interval** – governs how frequently the time service will adjust system time towards accuracy. AWS configures the update interval to occur once every two minutes.
- **NTP Server** – starting with the August 2018 release, AMIs will now use the Amazon Time Sync Service by default. This time service is accessible from any EC2 Region at the 169.254.169.123 endpoint. Additionally, the 0x9 flag indicates that the time service is acting as a client, and to use SpecialPollInterval to determine how frequently to check in with the configured time server.
- **Type** – "NTP" means that the service acts as a standalone NTP client instead of acting as part of a domain.
- **Enabled and InputProvider** – the time service is enabled and provides time to the operating system.
- **Special Poll Interval** – checks against the configured NTP Server every 900 seconds, or 15 minutes.

<table>
<thead>
<tr>
<th>Registry Path</th>
<th>Key Name</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>HKLM:\System \CurrentControlSet\services \w32time\Config</td>
<td>UpdateInterval</td>
<td>120</td>
</tr>
<tr>
<td>HKLM:\System \CurrentControlSet\services \w32time\Parameters</td>
<td>NtpServer</td>
<td>169.254.169.123,0x9</td>
</tr>
<tr>
<td>HKLM:\System \CurrentControlSet\services \w32time\Parameters</td>
<td>Type</td>
<td>NTP</td>
</tr>
<tr>
<td>HKLM:\System \CurrentControlSet\services \w32time\TimeProviders \NtpClient</td>
<td>Enabled</td>
<td>1</td>
</tr>
</tbody>
</table>

462
**Setting the Time**

### Registry Path

<table>
<thead>
<tr>
<th>Registry Path</th>
<th>Key Name</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>HKLM:\System\CurrentControlSet\services\w32time\TimeProviders\NtpClient</td>
<td>InputProvider</td>
<td>1</td>
</tr>
<tr>
<td>HKLM:\System\CurrentControlSet\services\w32time\TimeProviders\NtpClient</td>
<td>SpecialPollInterval</td>
<td>900</td>
</tr>
</tbody>
</table>

**Configuring Time Settings for Windows Server 2008 and later**

When you change the time on a Windows instance, you must ensure that the time persists through system restarts. Otherwise, when the instance restarts, it reverts back to using UTC time. For Windows Server 2008 and later, you can persist your time setting by adding a `RealTimeIsUniversal` registry key.

**To set the `RealTimeIsUniversal` registry key**

1. From the instance, open a Command Prompt window.
2. Use the following command to add the registry key:
   ```cmd
   reg add "HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\TimeZoneInformation" /v RealTimeIsUniversal /d 1 /t REG_DWORD /f
   ```
3. If you are using a Windows Server 2008 AMI (not Windows Server 2008 R2) that was created before February 22, 2013, we recommend updating to the latest AWS Windows AMI. If you are using an AMI running Windows Server 2008 R2 (not Windows Server 2008), you must verify that the Microsoft hotfix KB2922223 is installed. If this hotfix is not installed, we recommend updating to the latest AWS Windows AMI.
4. (Optional) Verify that the instance saved the key successfully using the following command:
   ```cmd
   reg query "HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\TimeZoneInformation" /s
   ```

   This command returns the subkeys for the `TimeZoneInformation` registry key. You should see the `RealTimeIsUniversal` key at the bottom of the list, similar to the following:

<table>
<thead>
<tr>
<th>Key Name</th>
<th>Type</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>REG_DWORD</td>
<td>0x1e0</td>
</tr>
<tr>
<td>DaylightBias</td>
<td>REG_DWORD</td>
<td>0xfffffffff4</td>
</tr>
<tr>
<td>DaylightName</td>
<td>REG_SZ</td>
<td>@tzres.dll,-211</td>
</tr>
<tr>
<td>DaylightStart</td>
<td>REG_BINARY</td>
<td>0000030002000200000000000000000000</td>
</tr>
<tr>
<td>StandardBias</td>
<td>REG_DWORD</td>
<td>0x0</td>
</tr>
<tr>
<td>StandardName</td>
<td>REG_SZ</td>
<td>@tzres.dll,-212</td>
</tr>
<tr>
<td>StandardStart</td>
<td>REG_BINARY</td>
<td>00000B0001000200000000000000000000</td>
</tr>
<tr>
<td>TimeZoneKeyName</td>
<td>REG_SZ</td>
<td>Pacific Standard Time</td>
</tr>
<tr>
<td>DynamicDaylightTimeDisabled</td>
<td>REG_DWORD</td>
<td>0x0</td>
</tr>
<tr>
<td>ActiveTimeBias</td>
<td>REG_DWORD</td>
<td>0x1a4</td>
</tr>
<tr>
<td>RealTimeIsUniversal</td>
<td>REG_DWORD</td>
<td>0x1</td>
</tr>
</tbody>
</table>

**Related Topics**

For more information about how the Windows operating system coordinates and manages time, including the addition of a leap second, see the following documentation:
Setting the Password

- How the Windows Time Service Works (Microsoft)
- W32tm (Microsoft)
- How the Windows Time service treats a leap second (Microsoft)
- The story around Leap Seconds and Windows: It's likely not Y2K (Microsoft)

Setting Passwords for Windows Instances

When you connect to a Windows instance, you must specify a user account and password that has permission to access the instance. The first time that you connect to an instance, you are prompted to specify the Administrator account and the default password.

With AWS Windows AMIs for Windows Server 2012 R2 and earlier, the EC2Config service (p. 405) generates the default password. With AWS Windows AMIs for Windows Server 2016 and later, EC2Launch (p. 395) generates the default password.

Note
With Windows Server 2016 and later, Password never expires is disabled for the local administrator. With Windows Server 2012 R2 and earlier, Password never expires is enabled for the local administrator.

Changing the Administrator Password After Connecting

When you connect to an instance the first time, we recommend that you change the Administrator password from its default value. Use the following procedure to change the Administrator password for a Windows instance.

Important
Store the new password in a safe place. You won't be able to retrieve the new password using the Amazon EC2 console. The console can only retrieve the default password. If you attempt to connect to the instance using the default password after changing it, you'll get a "Your credentials did not work" error.

To change the local Administrator password

1. Connect to the instance and open a command prompt.
2. Run the following command. If your new password includes special characters, ensure that you enclose the password in double quotes:
   ```
   net user Administrator "new_password"
   ```
3. Store the new password in a safe place.

Changing a Lost or Expired Password

If you lose your password or it expires, you can generate a new password. For password reset procedures, see Resetting a Lost or Expired Windows Administrator Password (p. 1058).

Adding Windows Components Using Installation Media

Windows Server operating systems include many optional components. Including all optional components in each Amazon EC2 Windows Server AMI is not practical. Instead, we provide you with installation media EBS snapshots that have the necessary files to configure or install components on your Windows instance.
To access and install the optional components, you must find the correct EBS snapshot for your version of Windows Server, create a volume from the snapshot, and attach the volume to your instance.

Before You Begin

Use the AWS Management Console or a command line tool to get the instance ID and Availability Zone of your instance. You must create your EBS volume in the same Availability Zone as your instance.

Adding Windows Components Using the Console

Use the following procedure to use the AWS Management Console to add Windows components to your instance.

To add Windows components to your instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Snapshots.
3. From the Filter bar, choose Public Snapshots.
4. Add the Owner filter and choose Amazon images.
5. Add the Description filter and type Windows.
6. Press Enter
7. Select the snapshot that matches your system architecture and language preference. For example, select Windows 2019 English Installation Media if your instance is running Windows Server 2019.
8. Choose Actions, Create Volume.
9. In the Create Volume dialog box, select the Availability Zone that matches your Windows instance, and then choose Create.
10. In the Volume Successfully Created message, choose the volume that you just created.
11. Choose Actions, Attach Volume.
12. In the Attach Volume dialog box, type the instance ID and the name of the device for the attachment, and choose Attach. If you need help with the device name, see Device Naming.
13. Connect to your instance and make the volume available. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804).

Important
Do not initialize the volume.
14. Open Control Panel, Programs and Features. Choose Turn Windows features on or off. If you are prompted for installation media, specify the EBS volume with the installation media.

Adding Windows Components Using the Tools for Windows PowerShell

Use the following procedure to use the Tools for Windows PowerShell to add Windows components to your instance.

To add Windows components to your instance using the Tools for Windows PowerShell

1. Use the Get-EC2Snapshot cmdlet with the Owner and description filters to get a list of the available installation media snapshots.

```powershell
PS C:\> Get-EC2Snapshot -Owner amazon -Filter @( Name="description"; Values="Windows*" )
```
2. In the output, note the ID of the snapshot that matches your system architecture and language preference. For example:

...  
DataEncryptionKeyId :
Description         : Windows 2019 English Installation Media
Encrypted           : False
KmsKeyId            :
OwnerAlias          : amazon
OwnerId             : 123456789012
Progress            : 100%
SnapshotId          : snap-22da283e
StartTime           : 10/25/2019 8:00:47 PM
State               : completed
StateMessage        :
Tags                :
VolumeId            : vol-be5eafcb
VolumeSize          : 6
...

3. Use the `New-EC2Volume` cmdlet to create a volume from the snapshot. Specify the same Availability Zone as your instance.

```
PS C:\> New-EC2Volume -AvailabilityZone us-east-1a -VolumeType gp2 -SnapshotId snap-22da283e
```

4. In the output, note the volume ID.

```
Attachments      : {}
AvailabilityZone : us-east-1a
CreateTime       : 4/18/2017 10:50:25 AM
Encrypted        : False
Iops             : 100
KmsKeyId         :
Size             : 6
SnapshotId       : snap-22da283e
State            : creating
Tags             :
VolumeId         : vol-06aa9e1fbf0b82ed1
VolumeType       : gp2
```

5. Use the `Add-EC2Volume` cmdlet to attach the volume to your instance.

```
PS C:\> Add-EC2Volume -InstanceId i-087711ddd0f98f9489 -VolumeId vol-06aa9e1fbf0b82ed1 -Device xvdh
```

6. Connect to your instance and make the volume available. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804).

**Important**
Do not initialize the volume.

7. Open Control Panel, Programs and Features. Choose Turn Windows features on or off. If you are prompted for installation media, specify the EBS volume with the installation media.

## Adding Windows Components Using the AWS CLI

Use the following procedure to use the AWS CLI to add Windows components to your instance.
To add Windows components to your instance using the AWS CLI

1. Use the `describe-snapshots` command with the `owner-ids` parameter and `description` filter to get a list of the available installation media snapshots.

```bash
aws ec2 describe-snapshots --owner-ids amazon --filters
Name=description,Values=Windows*
```

2. In the output, note the ID of the snapshot that matches your system architecture and language preference. For example:

```json
{
   "Snapshots": [
   ...
   {
      "OwnerAlias": "amazon",
      "Description": "Windows 2019 English Installation Media",
      "Encrypted": false,
      "VolumeId": "vol-be5eafcb",
      "State": "completed",
      "VolumeSize": 6,
      "Progress": "100%",
      "StartTime": "2019-10-25T20:00:47.000Z",
      "SnapshotId": "snap-22da283e",
      "OwnerId": "123456789012"
   },
   ...
   ]
}
```

3. Use the `create-volume` command to create a volume from the snapshot. Specify the same Availability Zone as your instance.

```bash
aws ec2 create-volume --snapshot-id snap-22da283e --volume-type gp2 --availability-zone us-east-1a
```

4. In the output, note the volume ID.

```json
{
   "AvailabilityZone": "us-east-1a",
   "Encrypted": false,
   "VolumeType": "gp2",
   "VolumeId": "vol-0c98b37f30bcb290",
   "State": "creating",
   "Iops": 100,
   "SnapshotId": "snap-22da283e",
   "CreateTime": "2017-04-18T10:33:10.940Z",
   "Size": 6
}
```

5. Use the `attach-volume` command to attach the volume to your instance.

```bash
aws ec2 attach-volume --volume-id vol-0c98b37f30bcb290 --instance-id i-01474ef662b89480 --device xvdg
```

6. Connect to your instance and make the volume available. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804).

**Important**
Do not initialize the volume.
Configuring a Secondary Private IPv4 Address for Your Windows Instance

You can specify multiple private IPv4 addresses for your instances. After you assign a secondary private IPv4 address to an instance, you must configure the operating system on the instance to recognize the secondary private IPv4 address.

Configuring the operating system on a Windows instance to recognize a secondary private IPv4 address requires the following:

- **Step 1: Configure Static IP Addressing on Your Windows Instance** (p. 468)
- **Step 2: Configure a Secondary Private IP Address for Your Windows Instance** (p. 470)
- **Step 3: Configure Applications to Use the Secondary Private IP Address** (p. 471)

**Note**
These instructions are based on Windows Server 2008 R2. The implementation of these steps may vary based on the operating system of the Windows instance.

**Prerequisites**
Before you begin, make sure you meet the following requirements:

- As a best practice, launch your Windows instances using the latest AMIs. If you are using an older Windows AMI, ensure that it has the Microsoft hot fix referenced in [http://support.microsoft.com/kb/2582281](http://support.microsoft.com/kb/2582281).
- After you launch your instance in your VPC, add a secondary private IP address. For more information, see [Assigning a Secondary Private IPv4 Address](p. 664).
- To allow Internet requests to your website after you complete the tasks in these steps, you must configure an Elastic IP address and associate it with the secondary private IP address. For more information, see [Associating an Elastic IP Address with the Secondary Private IPv4 Address](p. 666).

**Step 1: Configure Static IP Addressing on Your Windows Instance**

To enable your Windows instance to use multiple IP addresses, you must configure your instance to use static IP addressing rather than a DHCP server.

**Important**
When you configure static IP addressing on your instance, the IP address must match exactly what is shown in the console, CLI, or API. If you enter these IP addresses incorrectly, the instance could become unreachable.

**To configure static IP addressing on a Windows instance**

1. Connect to your instance.
2. Find the IP address, subnet mask, and default gateway addresses for the instance by performing the following steps:
   - At a Command Prompt window, run the following command:
ipconfig /all

Review the following section in your output, and note the IPv4 Address, Subnet Mask, Default Gateway, and DNS Servers values for the network interface.

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP address</td>
<td>The IPv4 address obtained in step 2 above.</td>
</tr>
<tr>
<td>Subnet mask</td>
<td>The subnet mask obtained in step 2 above.</td>
</tr>
<tr>
<td>Default gateway</td>
<td>The default gateway address obtained in step 2 above.</td>
</tr>
<tr>
<td>Preferred DNS server</td>
<td>The DNS server obtained in step 2 above.</td>
</tr>
<tr>
<td>Alternate DNS server</td>
<td>The alternate DNS server obtained in step 2 above. If an alternate DNS server was not listed, leave this field blank.</td>
</tr>
</tbody>
</table>

Important
If you set the IP address to any value other than the current IP address, you will lose connectivity to the instance.
You will lose RDP connectivity to the Windows instance for a few seconds while the instance converts from using DHCP to static addressing. The instance retains the same IP address information as before, but now this information is static and not managed by DHCP.

Step 2: Configure a Secondary Private IP Address for Your Windows Instance

After you have set up static IP addressing on your Windows instance, you are ready to prepare a second private IP address.

To configure a secondary IP address for a Windows instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances and select your instance.
3. On the Description tab, note the secondary IP address.
4. Connect to your instance.
5. On your Windows instance, choose Start, Control Panel.
6. Choose Network and Internet, Network and Sharing Center.
7. Select the network interface (Local Area Connection) and choose Properties.
10. In the TCP/IP Address dialog box, type the secondary private IP address for IP address. For Subnet mask, type the same subnet mask that you entered for the primary private IP address in Step 1: Configure Static IP Addressing on Your Windows Instance (p. 468), and then choose Add.
11. Verify the IP address settings and choose OK.
12. Choose **OK**, **Close**.

13. To confirm that the secondary IP address has been added to the operating system, at a command prompt, run the command `ipconfig /all`.

### Step 3: Configure Applications to Use the Secondary Private IP Address

You can configure any applications to use the secondary private IP address. For example, if your instance is running a website on IIS, you can configure IIS to use the secondary private IP address.

**To configure IIS to use the secondary private IP address**

1. Connect to your instance.
2. Open Internet Information Services (IIS) Manager.
3. In the **Connections** pane, expand **Sites**.
4. Open the context (right-click) menu for your website and choose **Edit Bindings**.
5. In the **Site Bindings** dialog box, for **Type**, choose **http**, **Edit**.
6. In the **Edit Site Binding** dialog box, for **IP address**, select the secondary private IP address. (By default, each website accepts HTTP requests from all IP addresses.)

7. Choose **OK**, **Close**.

### Configure a Secondary Elastic Network Interface

You can attach a second elastic network interface to the instance.
To configure a second network interface

1. Configure the static IP addressing for the primary elastic network interface as per the procedures above in Step 1: Configure Static IP Addressing on Your Windows Instance (p. 468).
2. Configure the static IP addressing for the secondary elastic network interface as per the same procedures.

Running Commands on Your Windows Instance at Launch

When you launch a Windows instance in Amazon EC2, you can pass user data to the instance. Instance user data is treated as opaque data; it is up to the instance to interpret it. For example, you can specify data to be used by automated configuration tasks or specify scripts that are run after the instance starts. User data is processed by EC2Config (p. 405) on Windows Server 2012 R2 and earlier and by EC2Launch (p. 395) on Windows Server 2016 and later.

Contents
- User Data Scripts (p. 472)
- User Data Execution (p. 473)
- User Data and the Console (p. 475)
- User Data and the Tools for Windows PowerShell (p. 476)

User Data Scripts

For EC2Config or EC2Launch to execute scripts, you must enclose the script within a special tag when you add it to user data. The tag you use depends on whether the commands run in a Command Prompt window (batch commands) or using Windows PowerShell.

If you specify both a batch script and a Windows PowerShell script, the batch script runs first and the Windows PowerShell script runs next, regardless of the order in which they appear in the instance user data.

Syntax for Batch Scripts

Specify a batch script using the `script` tag. Separate the commands using line breaks. For example:

```
<script>
  echo Current date and time >> %SystemRoot%\Temp\test.log
  echo %DATE% %TIME% >> %SystemRoot%\Temp\test.log
</script>
```

By default, the user data scripts are executed one time when you launch the instance. To execute the user data scripts every time you reboot or start the instance, add `<persist>true</persist>` to the user data.

```
<script>
  echo Current date and time >> %SystemRoot%\Temp\test.log
  echo %DATE% %TIME% >> %SystemRoot%\Temp\test.log
</script>
<persist>true</persist>
```
Syntax for Windows PowerShell Scripts

The AWS Windows AMIs include the AWS Tools for Windows PowerShell, so you can specify these cmdlets in user data. If you associate an IAM role with your instance, you don't need to specify credentials to the cmdlets, as applications that run on the instance use the role's credentials to access AWS resources (for example, Amazon S3 buckets).

Specify a Windows PowerShell script using the `powershell` tag. Separate the commands using line breaks. For example:

```powershell
$file = $env:SystemRoot + "\Temp" + (Get-Date).ToString("MM-dd-yy-hh-mm")
New-Item $file -ItemType file
</powershell>
```

By default, the user data scripts are executed one time when you launch the instance. To execute the user data scripts every time you reboot or start the instance, add `<persist>true</persist>` to the user data.

```powershell
$file = $env:SystemRoot + "\Temp" + (Get-Date).ToString("MM-dd-yy-hh-mm")
New-Item $file -ItemType file
</powershell>
<persist>true</persist>
```

Base64 Encoding

If you're using the Amazon EC2 API or a tool that does not perform base64 encoding of the user data, you must encode the user data yourself. If not, an error is logged about being unable to find `script` or `powershell` tags to execute. The following is an example that encodes using Windows PowerShell.

```powershell
#$UserData = [System.Convert]::ToBase64String([System.Text.Encoding]::ASCII.GetBytes($Script))
```

The following is an example that decodes using PowerShell.

```powershell
#$Script = [System.Text.Encoding]::UTF8.GetString([System.Convert]::FromBase64String($UserData))
```

For more information about base64 encoding, see http://tools.ietf.org/html/rfc4648.

User Data Execution

By default, all AWS Windows AMIs have user data execution enabled for the initial launch. You can specify that user data scripts are executed the next time the instance reboots or restarts. Alternatively, you can specify that user data scripts are executed every time the instance reboots or restarts.

User data scripts are executed from the local administrator account when a random password is generated. Otherwise, user data scripts are executed from the System account.

Instance Launch

Any scripts in the instance user data are executed during the initial launch of the instance. If the `persist` tag is found, user data execution is enabled for subsequent reboots or starts. The log files for EC2Launch and EC2Config contain the output from the standard output and standard error streams.
With EC2Launch, the log file is `C:\ProgramData\Amazon\EC2-Windows\Launch\Log\UserdataExecution.log`.

**Note**

The `C:\ProgramData` folder might be hidden. To view the folder, you must show hidden files and folders.

The following information is logged as user data is executed.

- Userdata execution begins — The start of user data execution
- `<persist>` tag was provided: true — If the persist tag is found
- Running userdata on every boot — If the persist tag is found
- `<powershell>` tag was provided: running powershell content — If the powershell tag is found
- `<script>` tag was provided: running script content — If the script tag is found
- Message: The output from user scripts — If user data scripts are executed, their output is logged

With EC2Config, the log file is `C:\Program Files\Amazon\Ec2ConfigService\Logs\Ec2Config.log`. The following information is logged as user data is executed.

- `Ec2HandleUserData`: Message: Start running user scripts — The start of user data execution
- `Ec2HandleUserData`: Message: Re-enabled userdata execution — If the persist tag is found
- `Ec2HandleUserData`: Message: Could not find `<persist>` and `</persist>` — If the persist tag is not found
- `Ec2HandleUserData`: Message: The output from user scripts — If user data scripts are executed, their output is logged

**Subsequent Reboots or Starts**

When you update instance user data, user data scripts are not executed automatically when you reboot or start the instance. However, you can enable user data execution so that user data scripts are executed one time when you reboot or start the instance or every time you reboot or start the instance.

If you choose the **Shutdown with Sysprep** option, user data scripts are executed when the instance is rebooted or restarted, even if you did not enable user data execution for subsequent reboots or starts.

**To enable user data execution on Windows Server 2016 or later (EC2Launch)**

1. Connect to your Windows instance.
2. Open a PowerShell command window and run the following command:
   ```bash
   C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeInstance.ps1 -Schedule
   ```
3. Disconnect from your Windows instance. To execute updated scripts next time the instance is started, stop the instance and update the user data. For more information, see [View and Update the Instance User Data](p. 475).

**To enable user data execution on Windows Server 2012 R2 and earlier (EC2Config)**

1. Connect to your Windows instance.
2. Open `C:\Program Files\Amazon\Ec2ConfigService\Ec2ConfigServiceSetting.exe`.
3. For **User Data**, select **Enable UserData execution for next service start**.
4. Disconnect from your Windows instance. To execute updated scripts next time the instance is started, stop the instance and update the user data. For more information, see [View and Update the Instance User Data](p. 475).
User Data and the Console

You can specify instance user data when you launch the instance. If the root volume of the instance is an EBS volume, you can also stop the instance and update its user data.

Specify Instance User Data at Launch

When you launch an instance, you specify the script in Advanced Details, User data on the Step 3: Configure Instance Details page of the Launch Instance wizard. The example in the following image creates a file in the Windows temporary folder, using the current date and time in the file name. When you include <persist>true</persist>, the script is executed every time you reboot or start the instance. When you select As text, the Amazon EC2 console performs the base64 encoding for you.

View and Update the Instance User Data

You can view the instance user data for any instance, and you can update the instance user data for a stopped instance.

To update the user data for an instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances. the instance.
3. Select the instance and choose Actions, Instance State, Stop.
   Warning
   When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.
4. When prompted for confirmation, choose Yes, Stop. It can take a few minutes for the instance to stop.
5. With the instance still selected, choose Actions, Instance Settings, View/Change User Data. You can’t change the user data if the instance is running, but you can view it.
6. In the View/Change User Data dialog box, update the user data, and then choose Save. To execute user data scripts every time you reboot or start the instance, add <persist>true</persist>, as shown in the following example:
Restart the instance. If you enabled user data execution for subsequent reboots or starts, the updated user data scripts are executed as part of the instance start process.

User Data and the Tools for Windows PowerShell

You can use the Tools for Windows PowerShell to specify, modify, and view the user data for your instance. For information about viewing user data from your instance using instance metadata, see Retrieve Instance User Data (p. 481). For information about user data and the AWS CLI, see User Data and the AWS CLI in the Amazon EC2 User Guide for Linux Instances.

Example: Specify Instance User Data at Launch

Create a text file with the instance user data. To execute user data scripts every time you reboot or start the instance, add `<persist>true</persist>`, as shown in the following example:

```xml
<powershell>
$file = $env:SystemRoot + "\Temp\" + (Get-Date).ToString("MM-dd-yyyy-hh-mm")
New-Item $file -ItemType file
</powershell>
<persist>true</persist>
```

To specify instance user data when you launch your instance, use the `New-EC2Instance` command. This command does not perform base64 encoding of the user data for you. Use the following commands to encode the user data in a text file named `script.txt`.

```powershell
PS C:\> $Script = Get-Content -Raw script.txt
PS C:\> $UserData = [System.Convert]::ToBase64String([System.Text.Encoding]::ASCII.GetBytes($Script))
```

Use the `-UserData` parameter to pass the user data to the `New-EC2Instance` command.

```powershell
PS C:\> New-EC2Instance -ImageId ami-abcd1234 -MinCount 1 -MaxCount 1 -InstanceType m3.medium -KeyName my-key-pair -SubnetId subnet-12345678 -SecurityGroupIds sg-1a2b3c4d
```
Example: Update Instance User Data for a Stopped Instance

You can modify the user data of a stopped instance using the `Edit-EC2InstanceAttribute` command.

Create a text file with the new script. Use the following commands to encode the user data in the text file named `new-script.txt`.

```powershell
PS C:\> $NewScript = Get-Content -Raw new-script.txt
PS C:\> $NewUserData = [System.Convert]::ToBase64String([System.Text.Encoding]::ASCII.GetBytes($NewScript))
```

Use the `-UserData` and `-Value` parameters to specify the user data.

```powershell
PS C:\> Edit-EC2InstanceAttribute -InstanceId i-1234567890abcdef0 -Attribute userData -Value $NewUserData
```

Example: View Instance User Data

To retrieve the user data for an instance, use the `Get-EC2InstanceAttribute` command.

```powershell
PS C:\> (Get-EC2InstanceAttribute -InstanceId i-1234567890abcdef0 -Attribute userData).UserData
```

The following is example output. Note that the user data is encoded.

```
PHBvd2Vyc2hlbGw +DQpSZW5hbWUtQ29tcHV0ZXIgLU5ld05hbWUgdXNlci1kYXRhLXRlc3QNCjwvcG93ZXJzaGVsbD4=
```

Use the following commands to store the encoded user data in a variable and then decode it.

```powershell
PS C:\> $UserData_encoded = (Get-EC2InstanceAttribute -InstanceId i-1234567890abcdef0 -Attribute userData).UserData
PS C:\> [System.Text.Encoding]::UTF8.GetString([System.Convert]::FromBase64String($UserData_encoded))
```

The following is example output.

```
<powershell>
$file = $env:SystemRoot + "\Temp\" + (Get-Date).ToString("MM-dd-yy-hh-mm")
New-Item $file -ItemType file
</powershell>
<persist>true</persist>
```

Example: Rename the Instance to Match the Tag Value

To read the tag value, rename the instance on first boot to match the tag value, and reboot, use the `Get-EC2Tag` command. To run this command successfully, you must have a role with `ec2:DescribeTags` permissions because tag information is unavailable in the metadata and must be retrieved by API call. For more information on how to attach a role to an instance, see Attaching an IAM Role to an Instance.

**Note**

This script fails on Windows Server versions prior to 2008.

```powershell
```
$nameValue = (get-ec2tag -filter @{Name="resource-id";Value=$instanceid},@{Name="key";Value="Name"}).Value
$_pattern = "^([a-zA-Z0-9-]{1,15}$|\[0-9]{1,15}$)
##Verify Name Value satisfies best practices for Windows hostnames
If ($nameValue -match $pattern)
    {Try
        {Rename-Computer -NewName $nameValue -Restart -ErrorAction Stop}
    Catch
        {$_Exception.Message
        Write-Output "Rename failed: $_Exception.Message"}}
Else
    {Throw "Provided name not a valid hostname. Please ensure Name value is between 1 and 15 characters in length and contains only alphanumeric or hyphen characters"}
</powershell>

Instance Metadata and User Data

Instance metadata is data about your instance that you can use to configure or manage the running instance. Instance metadata is divided into categories. For more information, see Instance Metadata Categories (p. 482).

Important
Although you can only access instance metadata and user data from within the instance itself, the data is not protected by cryptographic methods. Anyone who can access the instance can view its metadata. Therefore, you should take suitable precautions to protect sensitive data (such as long-lived encryption keys). You should not store sensitive data, such as passwords, as user data.

You can also use instance metadata to access user data that you specified when launching your instance. For example, you can specify parameters for configuring your instance, or attach a simple script. You can also use this data to build more generic AMIs that can be modified by configuration files supplied at launch time. For example, if you run web servers for various small businesses, they can all use the same AMI and retrieve their content from the Amazon S3 bucket you specify in the user data at launch. To add a new customer at any time, simply create a bucket for the customer, add their content, and launch your AMI. If you launch more than one instance at the same time, the user data is available to all instances in that reservation.

EC2 instances can also include dynamic data, such as an instance identity document that is generated when the instance is launched. For more information, see Dynamic Data Categories (p. 487).

Contents
- Retrieving Instance Metadata (p. 478)
- Working with Instance User Data (p. 480)
- Retrieving Dynamic Data (p. 481)
- Instance Metadata Categories (p. 482)
- Instance Identity Documents (p. 487)

Retrieving Instance Metadata

Because your instance metadata is available from your running instance, you do not need to use the Amazon EC2 console or the AWS CLI. This can be helpful when you're writing scripts to run from your instance. For example, you can access the local IP address of your instance from instance metadata to manage a connection to an external application.

To view all categories of instance metadata from within a running instance, use the following URI:

The IP address 169.254.169.254 is a link-local address and is valid only from the instance. For more information, see Link-local address on Wikipedia.

Note that you are not billed for HTTP requests used to retrieve instance metadata and user data.

You can use PowerShell cmdlets to retrieve the URI. For example, if you are running version 3.0 or later of PowerShell, use the following cmdlet:

```
```

If you don’t want to use PowerShell, you can install a third-party tool such as GNU Wget or cURL.

**Important**

If you do install a third-party tool on a Windows instance, ensure that you read the accompanying documentation carefully, as the method of calling the HTTP and the output format might be different from what is documented here.

All instance metadata is returned as text (content type text/plain). A request for a specific metadata resource returns the appropriate value, or a 404 - Not Found HTTP error code if the resource is not available.

A request for a general metadata resource (the URI ends with a /) returns a list of available resources, or a 404 - Not Found HTTP error code if there is no such resource. The list items are on separate lines, terminated by line feeds (ASCII 10).

**Examples of Retrieving Instance Metadata**

This example gets the available versions of the instance metadata. These versions do not necessarily correlate with an Amazon EC2 API version. The earlier versions are available to you in case you have scripts that rely on the structure and information present in a previous version.

```
1.0
2007-01-19
2007-03-01
2007-08-29
2007-10-10
2007-12-15
2008-02-01
2008-09-01
2009-04-04
2011-01-01
2011-05-01
2012-01-12
2014-02-25
2014-11-05
2015-10-20
2016-04-19
2016-06-30
2016-09-02
latest
```

This example gets the top-level metadata items. For more information, see Instance Metadata Categories (p. 482).

```
ami-id
ami-launch-index
ami-manifest-path
block-device-mapping/
hostname
iam/
```
instance-action
instance-id
instance-type
local-hostname
local-ipv4
mac
metrics/
network/
placement/
profile
public-hostname
public-ipv4
public-keys/
reservation-id
security-groups
services/

These examples get the value of some of the metadata items from the preceding example.

```powershell
ami-0abcdef1234567890

r-0efghjk987654321

ip-10-251-50-12.ec2.internal

e2-203-0-113-25.compute-1.amazonaws.com

This example gets the subnet ID for an instance.

macs/02:29:96:8f:6a:2d/subnet-id
subnet-be9b61d7

**Throttling**

We throttle queries to the instance metadata service on a per-instance basis, and we place limits on the number of simultaneous connections from an instance to the instance metadata service.

If you're using the instance metadata service to retrieve AWS security credentials, avoid querying for credentials during every transaction or concurrently from a high number of threads or processes, as this may lead to throttling. Instead, we recommend that you cache the credentials until they start approaching their expiry time.

If you're throttled while accessing the instance metadata service, retry your query with an exponential backoff strategy.

**Working with Instance User Data**

When working with instance user data, keep the following in mind:

- User data must be base64-encoded. The Amazon EC2 console can perform the base64 encoding for you or accept base64-encoded input.
- User data is limited to 16 KB, in raw form, before it is base64-encoded. The size of a string of length $n$ after base64-encoding is $\lceil n/3 \rceil \times 4$. 
• User data must be base64-decoded when you retrieve it. The data is decoded for you automatically if you retrieve it using instance metadata or the console.
• User data is treated as opaque data: what you give is what you get back. It is up to the instance to be able to interpret it.
• If you stop an instance, modify its user data, and start the instance, the updated user data is not executed automatically when you start the instance. However, you can configure settings so that updated user data scripts are executed one time when you start the instance or every time you reboot or start the instance.

Specify Instance User Data at Launch

You can specify user data when you launch an instance. You can specify that the user data is executed one time at launch, or every time you reboot or start the instance. For more information, see Running Commands on Your Windows Instance at Launch (p. 472).

Modify Instance User Data

You can modify user data for an instance in the stopped state if the root volume is an EBS volume. For more information, see View and Update the Instance User Data (p. 475).

Retrieve Instance User Data

To retrieve user data from within a running instance, use the following URI:

http://169.254.169.254/latest/user-data

A request for user data returns the data as it is (content type application/octet-stream).

This example returns user data that was provided as comma-separated text:

1234,john,reboot,true | 4512,richard, | 173,,,

This example returns user data that was provided as a script:

<powershell>
$file = $env:SystemRoot + "\Temp" + (Get-Date).ToString("MM-dd-yy-hh-mm")
New-Item $file -ItemType file
</powershell>
<persist>true</persist>

To retrieve user data for an instance from your own computer, see User Data and the Tools for Windows PowerShell (p. 476)

Retrieving Dynamic Data

To retrieve dynamic data from within a running instance, use the following URI:


This example shows how to retrieve the high-level instance identity categories:

document
rsa2048
pkcs7
For more information about dynamic data and examples of how to retrieve it, see Instance Identity Documents (p. 487).

**Instance Metadata Categories**

The following table lists the categories of instance metadata.

*Important*

Category names that are formatted in red text are placeholders for data that is unique to your instance; for example, `mac` represents the MAC address for the network interface. You must replace the placeholders with the actual values.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
<th>Version Introduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>ami-id</td>
<td>The AMI ID used to launch the instance.</td>
<td>1.0</td>
</tr>
<tr>
<td>ami-launch-index</td>
<td>If you started more than one instance at the same time, this value indicates the order in which the instance was launched. The value of the first instance launched is 0.</td>
<td>1.0</td>
</tr>
<tr>
<td>ami-manifest-path</td>
<td>The path to the AMI manifest file in Amazon S3. If you used an Amazon EBS-backed AMI to launch the instance, the returned result is unknown.</td>
<td>1.0</td>
</tr>
<tr>
<td>ancestor-ami-ids</td>
<td>The AMI IDs of any instances that were rebundled to create this AMI. This value will only exist if the AMI manifest file contained an ancestor-amis key.</td>
<td>2007-10-10</td>
</tr>
<tr>
<td>block-device-mapping/ami</td>
<td>The virtual device that contains the root/boot file system.</td>
<td>2007-12-15</td>
</tr>
<tr>
<td>block-device-mapping/ebs</td>
<td>The virtual devices associated with Amazon EBS volumes, if any are present. Amazon EBS volumes are only available in metadata if they were present at launch time or when the instance was last started. The <code>N</code> indicates the index of the Amazon EBS volume (such as ebs1 or ebs2).</td>
<td>2007-12-15</td>
</tr>
<tr>
<td>block-device-mapping/ephemeral</td>
<td>The virtual devices associated with non-NVMe instance store volumes, if any are present. The <code>N</code> indicates the index of each ephemeral volume.</td>
<td>2007-12-15</td>
</tr>
<tr>
<td>block-device-mapping/root</td>
<td>The virtual devices or partitions associated with the root devices, or partitions on the virtual device,</td>
<td>2007-12-15</td>
</tr>
<tr>
<td>Data</td>
<td>Description</td>
<td>Version Introduced</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>block-device-mapping/swap</td>
<td>The virtual devices associated with swap. Not always present.</td>
<td>2007-12-15</td>
</tr>
<tr>
<td>elastic-gpus/associations/elastic-gpu-id</td>
<td>If there is an Elastic GPU attached to the instance, contains a JSON string with information about the Elastic GPU, including its ID and connection information.</td>
<td>2016-11-30</td>
</tr>
<tr>
<td>elastic-inference/associations/eia-id</td>
<td>If there is an Elastic Inference accelerator attached to the instance, contains a JSON string with information about the Elastic Inference accelerator, including its ID and type.</td>
<td>2018-11-29</td>
</tr>
<tr>
<td>events/maintenance/history</td>
<td>If there are completed or canceled maintenance events for the instance, contains a JSON string with information about the events. For more information, see To view event history about completed or canceled events (p. 542).</td>
<td>2018-08-17</td>
</tr>
<tr>
<td>events/maintenance/scheduled</td>
<td>If there are active maintenance events for the instance, contains a JSON string with information about the events. For more information, see Viewing Scheduled Events (p. 540).</td>
<td>2018-08-17</td>
</tr>
<tr>
<td>hostname</td>
<td>The private IPv4 DNS hostname of the instance. In cases where multiple network interfaces are present, this refers to the eth0 device (the device for which the device number is 0).</td>
<td>1.0</td>
</tr>
<tr>
<td>iam/info</td>
<td>If there is an IAM role associated with the instance, contains information about the last time the instance profile was updated, including the instance's LastUpdated date, InstanceProfileArn, and InstanceProfileId. Otherwise, not present.</td>
<td>2012-01-12</td>
</tr>
<tr>
<td>iam/security-credentials/role-name</td>
<td>If there is an IAM role associated with the instance, role-name is the name of the role, and role-name contains the temporary security credentials associated with the role (for more information, see Retrieving Security Credentials from Instance Metadata (p. 647)). Otherwise, not present.</td>
<td>2012-01-12</td>
</tr>
<tr>
<td>Data</td>
<td>Description</td>
<td>Version Introduced</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>identity-credentials/ec2/info</td>
<td>[Reserved for internal use only] Information about the credentials that AWS uses to identify an instance to the rest of the Amazon EC2 infrastructure.</td>
<td>2018-05-23</td>
</tr>
<tr>
<td>identity-credentials/ec2/security-credentials/ec2-instance</td>
<td>[Reserved for internal use only] The credentials that AWS uses to identify an instance to the rest of the Amazon EC2 infrastructure.</td>
<td>2018-05-23</td>
</tr>
<tr>
<td>instance-action</td>
<td>Notifies the instance that it should reboot in preparation for bundling. Valid values: none</td>
<td>shutdown</td>
</tr>
<tr>
<td>instance-id</td>
<td>The ID of this instance.</td>
<td>1.0</td>
</tr>
<tr>
<td>instance-type</td>
<td>The type of instance. For more information, see Instance Types (p. 119).</td>
<td>2007-08-29</td>
</tr>
<tr>
<td>kernel-id</td>
<td>The ID of the kernel launched with this instance, if applicable.</td>
<td>2008-02-01</td>
</tr>
<tr>
<td>local-_hostname</td>
<td>The private IPv4 DNS hostname of the instance. In cases where multiple network interfaces are present, this refers to the eth0 device (the device for which the device number is 0).</td>
<td>2007-01-19</td>
</tr>
<tr>
<td>local-ipv4</td>
<td>The private IPv4 address of the instance. In cases where multiple network interfaces are present, this refers to the eth0 device (the device for which the device number is 0).</td>
<td>1.0</td>
</tr>
<tr>
<td>mac</td>
<td>The instance's media access control (MAC) address. In cases where multiple network interfaces are present, this refers to the eth0 device (the device for which the device number is 0).</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>metrics/vhostmd</td>
<td>Deprecated.</td>
<td>2011-05-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/device-number</td>
<td>The unique device number associated with that interface. The device number corresponds to the device name; for example, a device-number of 2 is for the eth2 device. This category corresponds to the DeviceIndex and device-index fields that are used by the Amazon EC2 API and the EC2 commands for the AWS CLI.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>Data</td>
<td>Description</td>
<td>Version Introduced</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/interface-id</td>
<td>The ID of the network interface.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/ipv4-assoc/p</td>
<td>The private IPv4 addresses that are associated with each public IP address and assigned to that interface.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>public-ip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>network/interfaces/macs/mac/ipv6s</td>
<td>The IPv6 addresses associated with the interface. Returned only for instances launched into a VPC.</td>
<td>2016-06-30</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/local-hostname</td>
<td>The interface's local hostname.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/local-ipv4s</td>
<td>The private IPv4 addresses associated with the interface.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/mac</td>
<td>The instance's MAC address.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/owner-id</td>
<td>The ID of the owner of the network interface. In multiple-interface environments, an interface can be attached by a third party, such as Elastic Load Balancing. Traffic on an interface is always billed to the interface owner.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/public-hostname</td>
<td>The interface's public DNS (IPv4). This category is only returned if the enableDnsHostnames attribute is set to true. For more information, see Using DNS with Your VPC.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/public-ipv4s</td>
<td>The Public IP address or Elastic IP addresses associated with the interface. There may be multiple IPv4 addresses on an instance.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/security-groups</td>
<td>Security groups to which the network interface belongs.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/security-group-ids</td>
<td>The IDs of the security groups to which the network interface belongs.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/subnet-id</td>
<td>The ID of the subnet in which the interface resides.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/subnet-ipv4-cidr-block</td>
<td>The IPv4 CIDR block of the subnet in which the interface resides.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/subnet-ipv6-cidr-block</td>
<td>The IPv6 CIDR block of the subnet in which the interface resides.</td>
<td>2016-06-30</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/vpc-id</td>
<td>The ID of the VPC in which the interface resides.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>Data</td>
<td>Description</td>
<td>Version Introduced</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/vpc-ipv4-cidr-block</td>
<td>The primary IPv4 CIDR block of the VPC.</td>
<td>2011-01-01</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/vpc-ipv4-cidr-blocks</td>
<td>The IPv4 CIDR blocks for the VPC.</td>
<td>2016-06-30</td>
</tr>
<tr>
<td>network/interfaces/macs/mac/vpc-ipv6-cidr-blocks</td>
<td>The IPv6 CIDR block of the VPC in which the interface resides.</td>
<td>2016-06-30</td>
</tr>
<tr>
<td>placement/availability-zone</td>
<td>The Availability Zone in which the instance launched.</td>
<td>2008-02-01</td>
</tr>
<tr>
<td>product-codes</td>
<td>Marketplace product codes associated with the instance, if any.</td>
<td>2007-03-01</td>
</tr>
<tr>
<td>public-hostname</td>
<td>The instance's public DNS. This category is only returned if the enableDnsHostnames attribute is set to true. For more information, see Using DNS with Your VPC in the Amazon VPC User Guide.</td>
<td>2007-01-19</td>
</tr>
<tr>
<td>public-ipv4</td>
<td>The public IPv4 address. If an Elastic IP address is associated with the instance, the value returned is the Elastic IP address.</td>
<td>2007-01-19</td>
</tr>
<tr>
<td>public-keys/0/openssh-key</td>
<td>Public key. Only available if supplied at instance launch time.</td>
<td>1.0</td>
</tr>
<tr>
<td>ramdisk-id</td>
<td>The ID of the RAM disk specified at launch time, if applicable.</td>
<td>2007-10-10</td>
</tr>
<tr>
<td>reservation-id</td>
<td>The ID of the reservation.</td>
<td>1.0</td>
</tr>
<tr>
<td>security-groups</td>
<td>The names of the security groups applied to the instance.</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>After launch, you can change the security groups of the instances.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Such changes are reflected here and in network/interfaces/macs/mac/security-groups.</td>
<td></td>
</tr>
<tr>
<td>services/domain</td>
<td>The domain for AWS resources for the region.</td>
<td>2014-02-25</td>
</tr>
<tr>
<td>services/partition</td>
<td>The partition that the resource is in. For standard AWS regions, the partition is aws. If you have resources in other partitions, the partition is aws-partitionname. For example, the partition for resources in the China (Beijing) region is aws-cn.</td>
<td>2015-10-20</td>
</tr>
</tbody>
</table>
Instance Metadata and User Data

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
<th>Version Introduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>spot/instance-action</td>
<td>The action (hibernate, stop, or terminate) and the approximate time, in UTC, when the action will occur. This item is present only if the Spot Instance has been marked for hibernate, stop, or terminate. For more information, see instance-action (p. 290).</td>
<td>2016-11-15</td>
</tr>
<tr>
<td>spot/termination-time</td>
<td>The approximate time, in UTC, that the operating system for your Spot Instance will receive the shutdown signal. This item is present and contains a time value (for example, 2015-01-05T18:02:00Z) only if the Spot Instance has been marked for termination by Amazon EC2. The termination-time item is not set to a time if you terminated the Spot Instance yourself. For more information, see termination-time (p. 290).</td>
<td>2014-11-05</td>
</tr>
</tbody>
</table>

Dynamic Data Categories

The following table lists the categories of dynamic data.

<table>
<thead>
<tr>
<th>Data</th>
<th>Description</th>
<th>Version introduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>fws/instance-monitoring</td>
<td>Value showing whether the customer has enabled detailed one-minute monitoring in CloudWatch. Valid values: enabled</td>
<td>2009-04-04</td>
</tr>
<tr>
<td>instance-identity/document</td>
<td>JSON containing instance attributes, such as instance-id, private IP address, etc. See Instance Identity Documents (p. 487).</td>
<td>2009-04-04</td>
</tr>
<tr>
<td>instance-identity/pkcs7</td>
<td>Used to verify the document's authenticity and content against the signature. See Instance Identity Documents (p. 487).</td>
<td>2009-04-04</td>
</tr>
<tr>
<td>instance-identity/signature</td>
<td>Data that can be used by other parties to verify its origin and authenticity. See Instance Identity Documents (p. 487).</td>
<td>2009-04-04</td>
</tr>
</tbody>
</table>

Instance Identity Documents

An instance identity document is a JSON file that describes an instance. The instance identity document is accompanied by a signature and a PKCS7 signature which can be used to verify the accuracy, origin, and authenticity of the information provided in the document.
The instance identity document is generated when the instance is launched, and exposed to the instance through instance metadata (p. 478). It validates the attributes of the instances, such as the instance size, instance type, operating system, and AMI.

**Important**
Due to the dynamic nature of instance identity documents and signatures, we recommend retrieving the instance identity document and signature regularly.

### Obtaining the Instance Identity Document and Signatures

To retrieve the instance identity document, use the following command from your running instance:

```
```

The following is example output:

```
privateIp          : 10.0.2.174
availabilityZone   : us-west-2a
developerProductCodes :
version            : 2010-08-31
instanceId         : i-1234567890abcdef0
billingProducts    : {bp-6ba54002}
instanceType       : m3.medium
architecture       : x86_64
accountId          : 123456789012
kernelId           :
ramdiskId          :
imageId            : ami-1562d075
region             : us-west-2
```

To retrieve the instance identity signature, use the following command from your running instance:

```
```

The following is example output:

```
dExamplesjNQhhJjan7pORLPLeSr7LJEFE4V2DhRGlyoYVBoUYrY9njyBCmHEaYaGhTS/AWY+LPx
lVSQURF5n0gwPNCuO6ICT0fNrm51H7w9ydaexamplejJw6XVWPxfuRkcN0TAAI4p4RtCAyq4ms
```

To retrieve the PKCS7 signature, use the following command from your running instance:

```
```

The following is example output:

```
MIICiTCCAfICCQD6m7oRW0uXOjANBkgqxhki9Gw0BAQUFADCBiDELMAkGA1UEBhMC
VVMxZzae2BgaNBvBrgtA1dBMBRAwDgYDVQQEwdfTWF0dGl1Mq8wDQYDVQQKEwZzbWF6
b24xPDASgNBvBAsTC01BSTSBdb25z2x1LMRwEAYDVQQQDEwluUXZ1Q21sYWMxHsAd
BgqXkXhGw0BCaEWE5vb251QGftYXpvi5bj2oWHehCMTMyeWND1MjA0NjIxWhcN
MTIvWNIyMjACOTiXXwCbicDELMakGA1UEBhMCVVMxZzae2BgaNBvBrgtA1dBMBRAwDgYD
VQQHdwdTWF0dGl1Mq8wDQYDVQQKEwZzbWF6b24xPDASgNBvBAsTC01BSTSBdb25z
b2x1LMRIwEAYDVQQQDEwluUXZ1Q21sYWMxHsAdBkgqxhki9Gw0BCaEWE5vb251QGft
```
Best Practices and Recommendations for SQL Server Clustering in EC2

SQL Always On clustering offers high availability without the requirement for shared storage. The list of practices in this topic, in addition to the prerequisites listed at Prerequisites, Restrictions, and Recommendations for Always On availability groups, can help you get the best results when operating a SQL Server Always On cluster on AWS. The practices listed in this topic also offer a method to gather logs.

Note
When nodes are deployed in different Availability Zones, or in different subnets within the same zone, they should be treated as a multi-subnet cluster. Keep this in mind as you apply best practices and when you address possible failure scenarios.

Contents
- Assigning IP Addresses (p. 489)
- Cluster Properties (p. 490)
- Cluster Quorum Votes and 50/50 Splits in a Multi-Site Cluster (p. 490)
- DNS Registration (p. 490)
- Elastic Network Adapters (ENAs) (p. 491)
- Multi-Site Clusters and EC2 Instance Placement (p. 491)
- Instance Type Selection (p. 491)
- Assigning Elastic Network Interfaces and IPs to the Instance (p. 491)
- Heartbeat Network (p. 492)
- Configuring the Network Adapter in the OS (p. 492)
- IPv6 (p. 492)
- Host Record TTL for SQL Availability Group Listeners (p. 492)
- Logging (p. 492)
- NetBIOS over TCP (p. 493)
- NetFT Virtual Adapter (p. 493)
- Setting Possible Owners (p. 493)
- Tuning the Failover Thresholds (p. 494)
- Witness Importance and Dynamic Quorum Architecture (p. 495)
- Troubleshooting (p. 495)

Assigning IP Addresses

Each cluster node should have one elastic network interface assigned that includes three private IP addresses on the subnet: a primary IP address, a cluster IP address, and an Availability Group IP address. The operating system (OS) should have the NIC configured for DHCP. It should not be set for a static IP address because the IP addresses for the cluster IP and Availability Group will be handled virtually in the Failover Cluster Manager. The NIC can be configured for a static IP as long as it is configured to only
use the primary IP of eth0. If the other IPs are assigned to the NIC, it can cause network drops for the instance during failover events.

When the network drops because the IPs are incorrectly assigned, or when there is a failover event or network failure, it is not uncommon to see the following event log entries at the time of failure.

```
Isatap interface isatap.{9468661C-0AEB-41BD-BB8C-1F85981D5482} is no longer active.
```

```
Isatap interface isatap.{9468661C-0AEB-41BD-BB8C-1F85981D5482} with address fe80::5efe:169.254.1.105 has been brought up.
```

Because these messages seem to describe network issues, it is easy to mistake the cause of the outage or failure as a network error. However, these errors describe a symptom, rather than cause, of the failure. ISATAP is a tunneling technology that uses IPv6 over IPv4. When the IPv4 connection fails, the ISATAP adapter also fails. When the network issues are resolved, these entries should no longer appear in the event logs. Alternately, you can eliminate network errors by safely disabling ISATAP with the following command.

```
netsh int ipv6 isatap set state disabled
```

When you run this command, the adapter is removed from Device Manager. This command should be run on all nodes. It does not impact the ability of the cluster to function. Instead, when the command has been run, ISATAP is no longer used. However, because this command might cause unknown impacts on other applications that leverage ISATAP, you should test it.

**Cluster Properties**

To see the complete cluster configuration, run the following PowerShell command.

```
Get-Cluster | Format-List -Property *
```

**Cluster Quorum Votes and 50/50 Splits in a Multi-Site Cluster**

To learn how the cluster quorum works and what to expect if a failure occurs, see Understanding Cluster and Pool Quorum.

**DNS Registration**

In Windows Server 2012, Failover Clustering, by default, attempts to register each DNS node under the cluster name. This is acceptable for applications that are aware the SQL target is configured for multi-site. However, when the client is not configured this way, it can result in timeouts, delays, and application errors due to attempts to connect to each individual node and failing on the inactive ones. To prevent these problems, the Cluster Resource parameter RegisterAllProvidersIp must be changed to 0. For more information, see RegisterAllProvidersIP Setting and Multi-subnet Clustered SQL + RegisterAllProvidersIP + SharePoint 2013.

The RegisterAllProvidersIp can be modified with the following PowerShell script.

```
Import-Module FailoverClusters
#cluster = (Get-ClusterResource | where {($_.ResourceType -eq "Network Name") -and ($_.OwnerGroup -ne "Cluster Group"))}.Name
Get-ClusterResource $cluster | Set-ClusterParameter RegisterAllProvidersIP 0
Get-ClusterResource $cluster | Set-ClusterParameter HostRecordTTL 300
Stop-ClusterResource $cluster
```
In addition to setting the Cluster Resource parameter to 0, you must ensure that the cluster has permissions to modify the DNS entry for your cluster name.

1. Log into the Domain Controller (DC) for the domain, or a server that hosts the forward lookup zone for the domain.
2. Launch the DNS Management Console and locate the A record for the cluster.
3. Right-click the A record and choose Properties.
5. Choose Add.
6. Choose Object Types..., select the box for Computers, and choose OK.
7. Enter the name of the cluster resource object and choose Check name and OK if resolve.
8. Select the check box for Full Control.
9. Choose OK.

**Elastic Network Adapters (ENAs)**

AWS has identified known issues with some clustering workloads running on ENA driver version 1.2.3. We recommend upgrading to version 1.5.0 or later and adjusting settings on the NIC in the OS. For the latest versions, see Amazon ENA Driver Versions. The first setting, which applies to all systems, increases Receive Buffers, which can be done with the following example PowerShell command.

```powershell
Set-NetAdapterAdvancedProperty -Name (Get-NetAdapter | Where-Object {$_._InterfaceDescription -like '*Elastic*').Name -DisplayName "Receive Buffers" -DisplayValue 8192
```

For instances with more than 16 vCPUs, we recommend preventing RSS from running on CPU 0.

Run the following command.

```powershell
Set-NetAdapterRss -name (Get-NetAdapter | Where-Object {$_._InterfaceDescription -like 'Elastic'}).Name -Baseprocessorgroup 0 -BaseProcessorNumber 1
```

**Multi-Site Clusters and EC2 Instance Placement**

Each cluster is considered a multi-site cluster. The EC2 service does not share IP addresses virtually. Each node must be in a unique subnet. Though not required, we recommend that each node also be in a unique Availability Zone (p. 5).

**Instance Type Selection**

The type of instance recommended for Windows Server Failover Clustering depends on the workload. For production workloads, we recommend instances that support EBS Optimization and Enhanced Networking.

**Assigning Elastic Network Interfaces and IPs to the Instance**

Each node in an EC2 cluster should have only one attached elastic network interface. The network interface should have a minimum of two assigned private IP addresses. However, for workloads that use Availability Groups, such as SQL Always On, you must include an additional IP address for each Availability Group. The primary IP address is used for accessing and managing the server; the secondary
IP address is used as the cluster IP address, and each additional IP address is assigned to Availability Groups, as needed.

**Heartbeat Network**

Some Microsoft documentation recommends using a dedicated heartbeat network. However, this recommendation is not applicable to EC2. With EC2, while you can assign and use a second elastic network interface for the heartbeat network, it uses the same infrastructure and shares bandwidth with the primary network interface. Therefore, traffic within the infrastructure cannot be prioritized, and cannot benefit from a dedicated network interface.

**Configuring the Network Adapter in the OS**

The NIC in the OS can keep using DHCP as long as the DNS servers that are being retrieved from the DHCP Options Set allow for the nodes to resolve each other. You can set the NIC to be configured statically. When completed, you then manually configure only the primary IP address for the elastic network interface. Failover Clustering manages and assigns additional IP addresses, as needed.

For all instance types, you can increase the MTU on the network adapter to 9001 to support Jumbo Frames. This configuration reduces fragmentation of packets wherever Jumbo Frames are supported. The following example shows how to use PowerShell to configure Jumbo Frames for an Elastic Network Adapter.

```
Get-NetAdapter | Set-NetAdapterAdvancedProperty -DisplayName "MTU" -DisplayValue 9001
```

**IPv6**

Microsoft does not recommend disabling IPv6 in a Windows Cluster. While Failover Clustering works in an IPv4-only environment, Microsoft tests clusters with IPv6 enabled. See Failover Clustering and IPv6 in Windows Server 2012 R2 for details.

**Host Record TTL for SQL Availability Group Listeners**

Set the host record TTL to 300 seconds instead of the default 20 minutes (1200 seconds). For legacy client comparability, set RegisterAllProvidersIP to 0 for SQL Availability Group Listeners. This is not required in all environments. These settings are important because some legacy client applications cannot use MultiSubnetFailover in their connection strings. See HostRecordTTL Setting for more information. When you change these settings, the Cluster Resource must be restarted. The following are example PowerShell scripts for changing the TTL and RegisterAllProvidersIP settings.

```
Get-ClusterResource yourListenerName | Set-ClusterParameter RegisterAllProvidersIP 0

Get-ClusterResource yourListenerName|Set-ClusterParameter HostRecordTTL 300

Stop-ClusterResource yourListenerName

Start-ClusterResource yourListenerName
```

**Logging**

The default logging level for the cluster log is 3. To increase the detail of log information, set the logging level to 5. See Set-ClusterLog for more information about the PowerShell cmdlet.
NetBIOS over TCP

On Windows Server 2012 R2, you can increase the speed of the failover process by disabling NetBIOS over TCP. This feature was removed from Windows Server 2016. You should test this procedure if you are using older operating systems in your environment. For more information, see Speeding Up Failover Tips-n-Tricks. The following is an example PowerShell command to disable NetBIOS over TCP.

```
Set-ClusterLog -Level 5

Get-ClusterResource "Cluster IP Address" | Set-ClusterParameter EnableNetBIOS 0
```

NetFT Virtual Adapter

For Windows Server versions earlier than 2016 and non-Hyper-V workloads, Microsoft recommends you enable the NetFT Virtual Adapter Performance Filter on the adapter in the OS. When you enable the NetFT Virtual Adapter, internal cluster traffic is routed directly to the NetFT Virtual Adapter. For more information, see NetFT Virtual Adapter Performance Filter. You can enable NetFT Virtual Adapter by selecting the check box in the NIC properties, or by using the following PowerShell command.

```
Get-NetAdapter | Set-NetAdapterBinding -ComponentID ms_netftflt -Enable $true
```

Setting Possible Owners

The Failover Cluster Manager can be configured so that each IP address specified on the Cluster Core Resources and Availability Group resources can be brought online only on the node to which the IP belongs. When the Failover Cluster Manager is not configured for this and a failure occurs, there will be some delay in failover as the cluster attempts to bring up the IPs on nodes that do not recognize the address. For more information, see SQL Server Manages Preferred and Possible Owner Properties for AlwaysOn Availability Group/Role.

Each resource in a cluster has a setting for Possible Owners. This setting tells the cluster which nodes are permitted to "online" a resource. Each node is running on a unique subnet in a VPC. Because EC2 cannot share IPs between instances, the IP resources in the cluster can be brought online only by specific nodes. By default, each IP address that is added to the cluster as a resource has every node listed as a Possible Owner. This does not result in failures. However, during expected and unexpected failures, you can see errors in the logs about conflicting IPs and failures to bring IPs online. These errors can be ignored. If you set the Possible Owner property, you can eliminate these errors entirely, and also prevent down time while the services are moved to another node.
Tuning the Failover Thresholds

In Server 2012 R2, the network thresholds for the failover heartbeat network default to high values. See Tuning Failover Cluster Network Thresholds for details. This potentially unreliable configuration (for clusters with some distance between them) was addressed in Server 2016 with an increase in the number of heartbeats. It was discovered that clusters would fail over due to very brief transient network issues. The heartbeat network is maintained with UDP 3343, which is traditionally far less reliable than TCP and more prone to incomplete conversations. Although there are low-latency connections between AWS Availability Zones, there are still geographic separations with a number of "hops" separating resources. Within an Availability Zone, there may be some distance between clusters unless the customer is using Placement Groups or Dedicated Hosts. As a result, there is a higher possibility for heartbeat failure with UDP than with TCP-based heartbeats.

The only time a cluster should fail over is when there is a legitimate outage, such as a service or node that experiences a hard failover, as opposed to a few UDP packets lost in transit. To ensure legitimate outages, we recommend that you adjust the thresholds to match, or even exceed, the settings for Server 2016 listed in Tuning Failover Cluster Network Thresholds. You can change the settings with the following PowerShell commands.

```
(get-cluster).SameSubnetThreshold = 10

(get-cluster).CrossSubnetThreshold = 20
```

When you set these values, unexpected failovers should be dramatically reduced. You can fine-tune these settings by increasing the delays between heartbeats. However, we recommend that you send the
heartbeats more frequently with greater thresholds. Setting these thresholds even higher ensures that failovers occur only for hard failover scenarios, with longer delays before failing over. You must decide how much down time is acceptable for your applications.

After increasing the `SameSubnetThreshold` or `CrossSubnetThreshold`, we recommend that you increase the `RouteHistoryLength` to double the higher of the two values. This ensures that there is sufficient logging for troubleshooting. You can set the `RouteHistoryLength` with the following PowerShell command.

```powershell
(Get-Cluster).RouteHistoryLength = 20
```

### Witness Importance and Dynamic Quorum Architecture

There is a difference between Disk Witness and File Share Witness. Disk Witness keeps a backup of the cluster database while File Share Witness does not. Both add a vote to the cluster (p. 490). You can use Disk Witness if you use iSCSI-based storage. For more about witness options, see File Share witness vs Disk witness for local clusters.

### Troubleshooting

If you experience unexpected failovers, first make sure that you are not experiencing networking, service, or infrastructure issues.

1. Check that your nodes are not experiencing network-related issues.
2. Check driver updates. If you are using outdated drivers on your instance, you should update them. Updating your drivers might address bugs and stability issues that might be present in your currently installed version.
3. Check for any possible resource bottlenecks that could cause an instance to become unresponsive, such as CPU and disk I/O. If the node cannot service requests, it may appear to be down by the cluster service.

### Upgrading an Amazon EC2 Windows Instance to a Newer Version of Windows Server

There are two methods to upgrade an earlier version of Windows Server running on an instance: in-place upgrade and migration (also called side-by-side upgrade). An in-place upgrade upgrades the operating system files while your personal settings and files are intact. A migration involves capturing settings, configurations, and data and porting these to a newer operating system on a fresh Amazon EC2 instance.

Microsoft has traditionally recommended migrating to a newer version of Windows Server instead of upgrading. Migrating can result in fewer upgrade errors or issues, but can take longer than an in-place upgrade because of the need to provision a new instance, plan for and port applications, and adjust configurations settings on the new instance. An in-place upgrade can be faster, but software incompatibilities can produce errors.

### Contents

- Performing a Server Migration (p. 496)
- Performing an In-Place Upgrade (p. 496)
- Performing an Automated Upgrade (p. 500)
- Migrating to Latest Generation Instance Types (p. 506)
- Windows to Linux Replatforming Assistant for Microsoft SQL Server Databases (p. 511)
Performing a Server Migration

Migrating involves capturing settings, configurations, and data and porting these to a newer operating system on separate hardware. After validation, the migrated system can be promoted to production. You can migrate instances by launching a new instance from an AMI of the new operating system. You can streamline the process further by using AWS CloudFormation and AWS Systems Manager to automatically apply settings and configurations to the new system with little manual work.

To migrate your server

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose AMIs, Owned by me, and Public images.
3. For Search, add the following filters and press Enter. The Search values are case-sensitive.
   - Owner: Amazon images
4. Launch a new instance from an AMI.
5. Log on to the new instance and install all updates.
6. Perform an application installation and configuration changes.
7. Test the server.
8. When validated, promote the server to production.

Performing an In-Place Upgrade

Before you perform an in-place upgrade, you must determine which network drivers the instance is running. PV network drivers enable you to access your instance using Remote Desktop. Starting with Windows Server 2008 R2, instances use either AWS PV, Intel Network Adapter, or the Enhanced Networking drivers. Instances with Windows Server 2003 and Windows Server 2008 use Citrix PV drivers. For more information, see Paravirtual Drivers for Windows Instances (p. 431).

Automated Upgrades

For steps on how to use AWS Systems Manager to automate the upgrade of your Windows Server 2008 R2 to Server 2012 R2 or from SQL Server 2008 R2 on Windows Server 2012 R2 to SQL Server 2016, see Upgrade Your End of Support Microsoft 2008 Workloads in AWS with Ease.

Before You Begin an In-Place Upgrade

Complete the following tasks and note the following important details before you begin your in-place upgrade.

- Read the Microsoft documentation to understand the upgrade requirements, known issues, and restrictions. Also review the official instructions for upgrading.
  - Upgrading to Windows Server 2008
  - Upgrading to Windows Server 2008 R2
  - Upgrade Options for Windows Server 2012
  - Upgrade Options for Windows Server 2012 R2
  - Upgrade and conversion options for Windows Server 2016
• Upgrade and conversion options for Windows Server 2019
• Windows Server Upgrade Center

We recommend performing an operating system upgrade on instances with at least 2 vCPUs and 4GB of RAM. If needed, you can change the instance to a larger size of the same type (t2.small to t2.large, for example), perform the upgrade, and then resize it back to the original size. If you are required to retain the instance size, you can monitor the progress using the Instance Console Screenshot (p. 1051). For more information, see Changing the Instance Type (p. 188).

• Verify that the root volume on your Windows instance has enough free disk space. The Windows Setup process might not warn you of insufficient disk space. For information about how much disk space is required to upgrade a specific operating system, see the Microsoft documentation. If the volume does not have enough space, it can be expanded. For more information, see Amazon EBS Elastic Volumes (p. 855).

• Determine your upgrade path. You must upgrade the operating system to the same architecture. For example, you must upgrade a 32-bit system to a 32-bit system. Windows Server 2008 R2 and later are 64-bit only.

• Disable antivirus and anti-spyware software and firewalls. These types of software can conflict with the upgrade process. Re-enable antivirus and anti-spyware software and firewalls after the upgrade completes.

• Update to the latest drivers as described in Migrating to Latest Generation Instance Types (p. 506).

• The Upgrade Helper Service only supports instances running Citrix PV drivers. If the instance is running Red Hat drivers, you must manually upgrade those drivers (p. 436) first.

Upgrade an Instance In-Place with AWS PV, Intel Network Adapter, or the Enhanced Networking Drivers

Use the following procedure to upgrade a Windows Server instance using the AWS PV, Intel Network Adapter, or the Enhanced Networking network drivers.

To perform the in-place upgrade

1. Create an AMI of the system you plan to upgrade for either backup or testing purposes. You can then perform the upgrade on the copy to simulate a test environment. If the upgrade completes, you can switch traffic to this instance with little downtime. If the upgrade fails, you can revert to the backup. For more information, see Creating a Custom Windows AMI (p. 64).

2. Ensure that your Windows Server instance is using the latest network drivers. See Upgrading PV Drivers on Your Windows Instances (p. 436) for information on upgrading your AWS PV driver.

3. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.

4. In the navigation pane, choose Instances. Locate the instance. Make a note of the instance ID and Availability Zone for the instance. You need this information later in this procedure.

5. If you are upgrading from Windows Server 2012 or 2012 R2 to Windows Server 2016 or 2019, do the following on your instance before proceeding:
   a. Uninstall the EC2Config service. For more information, see Stopping, Restarting, Deleting, or Uninstalling EC2Config (p. 407).
   b. Install the EC2Launch service. For more information, see Installing the Latest Version of EC2Launch (p. 396).
   c. Install the AWS Systems Manager SSM Agent. For more information, see Working with SSM Agent in the AWS Systems Manager User Guide.

6. Create a new volume from a Windows Server installation media snapshot.
   a. In the navigation pane, choose Snapshots, Public Snapshots.
Performing an In-Place Upgrade

b. Add the **Owner** filter and choose **Amazon images**.

c. Add the **Description** filter and enter **Windows**. Press Enter.

d. Select the snapshot that matches the system architecture and language preference you are upgrading to. For example, select **Windows 2019 English Installation Media** to upgrade to Windows Server 2019.

e. Choose **Actions, Create Volume**.

f. In the **Create Volume** dialog box, choose the Availability Zone that matches your Windows instance, and choose **Create**.

7. In the **Volume Successfully Created** message, choose the volume that you just created.

8. Choose **Actions, Attach Volume**.

9. In the **Attach Volume** dialog box, enter the instance ID and choose **Attach**.

10. Begin the upgrade by using Windows PowerShell to open the installation media volume you attached to the instance.

   a. If you are upgrading to Windows Server 2016 or later, run the following:

   ```bash
   ./setup.exe /auto upgrade
   ```

   If you are upgrading to an earlier version of Windows Server, run the following:

   ```bash
   Sources/setup.exe
   ```

   b. For **Select the operating system you want to install**, select the full installation SKU for your Windows Server instance, and choose **Next**.

   c. For **Which type of installation do you want?**, choose **Upgrade**.

   d. Complete the wizard.

Windows Server Setup copies and processes files. After several minutes, your Remote Desktop session closes. The time it takes to upgrade depends on the number of applications and server roles running on your Windows Server instance. The upgrade process could take as little as 40 minutes or several hours. The instance fails status check 1 of 2 during the upgrade process. When the upgrade completes, both status checks pass. You can check the system log for console output or use Amazon CloudWatch metrics for disk and CPU activity to determine whether the upgrade is progressing.

**Note**

If upgrading to Windows Server 2019, after the upgrade is complete you can change the desktop background manually to remove the previous operating system name if desired.

If the instance has not passed both status checks after several hours, see **Troubleshooting an Upgrade** (p. 518).

### Upgrade an Instance In-Place with Citrix PV Drivers

Citrix PV drivers are used in Windows Server 2003 and 2008. There is a known issue during the upgrade process where Windows Setup removes portions of the Citrix PV drivers that enable you to connect to the instance by using Remote Desktop. To avoid this problem, the following procedure describes how to use the Upgrade Helper Service during your in-place upgrade.

#### Using the Upgrade Helper Service

You must run the Upgrade Helper Service before you start the upgrade. After you run it, the utility creates a Windows service that executes during the post-upgrade steps to correct the driver state. The executable is written in C# and can run on .NET Framework versions 2.0 through 4.0.
When you run Upgrade Helper Service on the system before the upgrade, it performs the following tasks:

- Creates a new Windows service named UpgradeHelperService.
- Verifies that the Citrix PV drivers are installed.
- Checks for unsigned boot critical drivers and presents a warning if any are found. Unsigned boot critical drivers could cause system failure after the upgrade if the drivers are not compatible with the newer Windows Server version.

When you run Upgrade Helper Service on the system after the upgrade, it performs the following tasks:

- Enables the RealTimeIsUniversal registry key for the correct time synchronization.
- Restores the missing PV driver by executing the following command:

  ```
  pnputil -i -a "C:\Program Files (x86)\Citrix\XenTools\*.inf"
  ```

- Installs the missing device by executing the following command:

  ```
  C:\Temp\EC2DriverUtils.exe install "C:\Program Files (x86)\Citrix\XenTools\xevtchn.inf"
  ROOT\XENEVTCHN
  ```

- Automatically removes UpgradeHelperService when complete.

### Performing the Upgrade on Instances Running Citrix PV Drivers

To complete the upgrade, you must attach the installation media volume to your EC2 instance and use UpgradeHelperService.exe.

#### To upgrade a Windows Server instance running Citrix PV drivers

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances** and locate the instance. Make a note of the instance ID and Availability Zone for the instance. You need this information later in this procedure.
3. Create a new volume from a Windows Server installation media snapshot.
   
   a. In the navigation pane, choose **Snapshots**, **Public Snapshots**.
   b. Add the **Owner** filter and choose **Amazon images**.
   c. Add the **Description** filter and enter **Windows**. Press Enter.
   d. Select the snapshot that matches the system architecture of your instance. For example, **Windows 2008 64-bit Installation Media**.
   e. Choose **Actions**, **Create Volume**.
   f. In the **Create Volume** dialog box, select the Availability Zone that matches your Windows instance, and choose **Create**.
4. In the **Volume Successfully Created** dialog box, choose the volume that you just created.
5. Choose **Actions**, **Attach Volume**.
6. In the **Attach Volume** dialog box, enter the instance ID and choose **Attach**.
7. On your Windows instance, on the C:\ drive, create a folder named temp.

   **Important**
   This folder must be available in the same location after the upgrade. Creating the folder in a Windows system folder or a user profile folder, such as the desktop, can cause the upgrade to fail.
8. Download OSUpgrade.zip and extract the files into the C:\temp folder.
9. Run `C:\temp\UpgradeHelperService.exe` and review the `C:\temp\Log.txt` file for any warnings.

10. Use Knowledge Base article 950376 from Microsoft to uninstall PowerShell from a Windows 2003 instance.

11. Begin the upgrade by using Windows Explorer to open the installation media volume that you attached to the instance.

12. Run the `Sources\Setup.exe` file.

13. For Select the operating system you want to install, select the full installation SKU for your Windows Server instance, and then choose Next.

14. For Which type of installation do you want?, choose Upgrade.

15. Complete the wizard.

Windows Server Setup copies and processes files. After several minutes, your Remote Desktop session closes. The time it takes to upgrade depends on the number of applications and server roles running on your Windows Server instance. The upgrade process could take as little as 40 minutes or several hours. The instance fails status check 1 of 2 during the upgrade process. When the upgrade completes, both status checks pass. You can check the system log for console output or use Amazon CloudWatch metrics for disk and CPU activity to determine whether the upgrade is progressing.

**Post Upgrade Tasks**

1. Log in to the instance to initiate an upgrade for the .NET Framework and reboot the system when prompted.

2. Install the latest version of the EC2Config service. For more information, see Installing the Latest Version of EC2Config (p. 406).

3. Install Microsoft hotfix KB2800213.

4. Install Microsoft hotfix KB2922223.

5. If you upgraded to Windows Server 2012 R2, we recommend that you upgrade the PV drivers to AWS PV drivers. For more information, see Windows Server 2012 R2.

6. Re-enable antivirus and anti-spyware software and firewalls.

**Performing an Automated Upgrade**

You can perform an automated upgrade on your Windows Server 2008 R2 and SQL Server 2008 R2 with Service Pack 3 instances on AWS with AWS Systems Manager SSM documents.

The Systems Manager Automation documents provide two upgrade paths:

- Windows Server 2008 R2 to Windows Server 2012 R2 using the SSM document for Automation named AWSEC2-CloneInstanceAndUpgradeWindows
- SQL Server 2008 R2 on Windows Server 2012 R2 to SQL Server 2016 using the SSM document for Automation named AWSEC2-CloneInstanceAndUpgradeSQLServer

**Contents**

- Related Services (p. 501)
- Prerequisites (p. 501)
- Upgrade Paths (p. 502)
- Steps for Performing an Automated Upgrade (p. 503)
Related Services

The following AWS services are used in the automated upgrade process:

- **AWS Systems Manager.** AWS Systems Manager is a powerful, unified interface for centrally managing your AWS resources. For more information, see the [AWS Systems Manager User Guide](https://docs.aws.amazon.com/systems-manager/latest/userguide/).

- **AWS Systems Manager Agent (SSM Agent).** This is Amazon software that can be installed and configured on an Amazon EC2 instance, an on-premises server, or a virtual machine (VM). SSM Agent makes it possible for Systems Manager to update, manage, and configure these resources. The agent processes requests from the Systems Manager service in the AWS Cloud, and then runs them as specified in the request. For more information, see [Working with SSM Agent](https://docs.aws.amazon.com/systems-manager/latest/userguide/) in the [AWS Systems Manager User Guide](https://docs.aws.amazon.com/systems-manager/latest/userguide/).

- **AWS Systems Manager SSM documents.** An SSM document defines the actions that Systems Manager performs on your managed instances. SSM documents use JavaScript Object Notation (JSON) or YAML, and they include steps and parameters that you specify. This topic uses two Systems Manager SSM documents for Automation. For more information, see [AWS Systems Manager Documents](https://docs.aws.amazon.com/systems-manager/latest/userguide/) in the [AWS Systems Manager User Guide](https://docs.aws.amazon.com/systems-manager/latest/userguide/).

Prerequisites

In order to automate your upgrade with AWS Systems Manager SSM documents, you must perform the following tasks:

- **Create an IAM role with the specified IAM policies (p. 501)** to allow Systems Manager to perform automation tasks on your Amazon EC2 instances and verify that you meet the prerequisites to use Systems Manager.

- **Select the option for how you want the automation to be executed (p. 501).** The options for execution are Simple execution, Rate control, Multi-account and Region, and Manual execution.

Create IAM Role with Specified Permissions

For steps on how to create an IAM role in order to allow AWS Systems Manager to access resources on your behalf, see [Creating a Role to Delegate Permissions to an AWS Service](https://docs.aws.amazon.com/IAM/latest/UserGuide/). This topic also contains information on how to verify that your account meets the prerequisites to use Systems Manager.

Select Execution Option

When you select Automation on the Systems Manager console, select Execute. After you select an SSM document, you are then prompted to choose an automation execution option. You choose from the following options. In the steps for the paths provided later in this topic, we use the Simple execution option.

Simple Execution

Choose this option if you want to update a single instance but do not want to go through each automation step to audit the results. This option is explained in further detail in the upgrade steps that follow.

Rate control

Choose this option if you want to apply the upgrade to more than one instance. You define the following settings.

- **Parameter**

  This setting, which is also set in Multi-Account and Region settings, defines how your automation branches out.
• Targets
Select the target to which you want to apply the automation. This setting is also set in Multi-Account and Region settings.

• Parameter Values
Use the values defined in the automation document parameters.

• Resource Group
In AWS, a resource is an entity you can work with. Examples include Amazon EC2 instances, AWS CloudFormation stacks, or Amazon S3 buckets. If you work with multiple resources, it may be useful to manage them as a group as opposed to moving from one AWS service to another for every task. In some cases, you may want to manage large numbers of related resources, such as EC2 instances that make up an application layer. In this case, you will likely need to perform bulk actions on these resources at one time.

• Tags
Tags help you categorize your AWS resources in different ways, for example, by purpose, owner, or environment. This categorization is useful when you have many resources of the same type. You can quickly identify a specific resource using the assigned tags.

• Rate Control
Rate Control is also set in Multi-Account and Region settings. When you set the rate control parameters, you define how many of your fleet to apply the automation to, either by target count or by percentage of the fleet.

Multi-Account and Region
In addition to the parameters specified under Rate Control that are also used in the Multi-Account and Region settings, there are two additional settings:

• Accounts and organizational units (OUs)
Specify multiple accounts on which you want to run the automation.

• AWS Regions
Specify multiple AWS Regions where you want to run the automation.

Manual Execution
This option is similar to Simple execution, but allows you to step through each automation step and audit the results.

Upgrade Paths
There are two upgrade paths, which use two different AWS Systems Manager Automation documents.

• AWSEC2-CloneInstanceAndUpgradeWindows. This script creates an Amazon Machine Image (AMI) from a Windows Server 2008 R2 instance in your account and upgrades this AMI to Windows Server 2012 R2. This multi-step process can take up to two hours to complete.

In this workflow, the automation creates an AMI from the instance and then launches the new AMI in the VPC and subnet you provide. The automation workflow performs an in-place upgrade from Windows Server 2008 R2 to Windows Server 2012 R2. It also updates or installs the AWS drivers required by the upgraded instance. After the upgrade is complete, the workflow creates a new AMI and terminates the upgraded instance.
Performing an Automated Upgrade

- **AWSEC2-CloneInstanceAndUpgradeSQLServer.** This script creates an AMI from an Amazon EC2 instance running SQL Server 2008 R2 SP3 in your account, and then upgrades the AMI to SQL Server 2016 SP2. This multi-step process can take up to two hours to complete.

In this workflow, the automation creates an AMI from the instance and then launches the new AMI in the subnet you provide. The automation then performs an in-place upgrade of SQL Server 2008 R2 to SQL Server 2016 SP2. After the upgrade is complete, the automation creates a new AMI before terminating the upgraded instance.

There are two AMIs included in the automated upgrade process:

- **Current running instance.** The first AMI is the current running instance, which is not upgraded. This AMI is used to launch another instance to run the in-place upgrade. When the process is complete, this AMI is deleted from your account, unless you specifically request to keep the original instance. This setting is handled by the parameter `KeepPreUpgradeImageBackUp` (default value is `false`, which means the AMI is deleted by default).

- **Upgraded AMI.** This AMI is the outcome of the automation process. The second AMI includes SQL Server 2016 SP2 instead of SQL Server 2008 R2.

The final result is one AMI, which is the upgraded instance of the AMI.

When the upgrade is complete, you can test your application functionality by launching the new AMI in your VPC. After testing, and before you perform another upgrade, schedule application downtime before completely switching to the upgraded instance.

### Steps for Performing an Automated Upgrade

**Upgrade Paths**

- Upgrade Windows 2008 R2 to 2012 R2 (p. 503)
- Upgrade SQL Server 2008 R2 to SQL Server 2016 (p. 504)

**Upgrade Windows 2008 R2 to 2012 R2**

This upgrade path requires additional prerequisites to work successfully. These prerequisites can be found in the automation document details for **AWSEC2-CloneInstanceAndUpgradeWindows** in the AWS Systems Manager User Guide.

After you have verified the additional prerequisite tasks, follow these steps to upgrade your Windows 2008 R2 instance to Windows 2012 R2 by using the automation document on AWS Systems Manager.

1. Open Systems Manager from the **AWS Management Console**.
2. From the left navigation pane, choose **Automation**.
3. Choose **Execute automation**.
4. Search for the automation document called **AWSEC2-CloneInstanceAndUpgradeWindows**.
5. When the document name appears, select it. When you select it, the document details appear.
6. Select **Next** to input the parameters for this document. Leave **Simple execution** selected at the top of the page.
7. Enter the requested parameters based on the following guidance.

- **InstanceID**

  **Type:** String

  (Required) The instance running Windows Server 2012 R2.
Performing an Automated Upgrade

- **InstanceProfile**
  
  **Type:** String
  
  (Required) The IAM instance profile. This is the IAM role used to perform the Systems Manager automation against the Amazon EC2 instance and AWS AMIs. For more information, see Create an IAM Instance Profile for Systems Manager in the AWS Systems Manager User Guide.

- **SubnetId**
  
  **Type:** String
  
  (Required) This is the subnet for the upgrade process and where your source EC2 instance resides. Verify that the subnet has outbound connectivity to AWS services, including Amazon S3, and also to Microsoft (in order to download patches).

- **KeepPreUpgradedBackUp**
  
  **Type:** String
  
  (Optional) If this parameter is set to `true`, the automation retains the image created from the instance. The default setting is `false`.

- **RebootInstanceBeforeTakingImage**
  
  **Type:** String
  
  (Optional) The default is `false` (no reboot). If this parameter is set to `true`, Systems Manager reboots the instance before creating an AMI for the upgrade.

8. After you have entered the parameters, select **Execute**. When the automation begins, you can monitor the execution progress.

9. When the automation completes, you will see the AMI ID. You can launch the AMI to verify that the Windows OS is upgraded.

**Note**

It is not necessary for the automation to run all of the steps. The steps are conditional based on the behavior of the automation and instance. Systems Manager might skip some steps that are not required. Additionally, some steps may time out. Systems Manager attempts to upgrade and install all of the latest patches. Sometimes, however, patches time out based on a definable timeout setting for the given step. When this happens, the Systems Manager automation continues to the next step to ensure that the internal OS is upgraded to Windows Server 2012 R2.

10. After the automation completes, you can launch an Amazon EC2 instance using the AMI ID to review your upgrade. For more information about how to create an Amazon EC2 instance from an AWS AMI, see How do I launch an EC2 instance from a custom Amazon Machine Image (AMI)?

### Upgrade SQL Server 2008 R2 to SQL Server 2016

This upgrade path requires additional prerequisites to work successfully. These prerequisites can be found in the automation document details for AWSEC2-CloneInstanceAndUpgradeSQLServer in the AWS Systems Manager User Guide.

After you have verified the additional prerequisite tasks, follow these steps to upgrade your SQL Server 2008 R2 database engine to SQL Server 2016 using the automation document on AWS Systems Manager.

1. If you haven't already, download the SQL Server 2016 .iso file and mount it to the source server.
2. After the .iso file is mounted, copy all of the component files and place them on any volume of your choice.
3. Take an EBS snapshot of the volume and copy the snapshot ID onto a clipboard for later use. For more information about creating an EBS snapshot, see Creating an EBS Snapshot in the Amazon Elastic Compute Cloud User Guide.

4. Attach the instance profile to the EC2 source instance. This allows Systems Manager to communicate with the EC2 instance and run commands on it after it is added to the AWS Systems Manager service. For this example, we named the role SSM-EC2-Profile-Role with the AmazonSSMManagedInstanceCore policy attached to the role. See Create an IAM Instance Profile for Systems Manager in the AWS Systems Manager User Guide.

5. In the AWS Systems Manager console, in the left navigation pane, choose Managed Instances. Verify that your EC2 instance is in the list of managed instance. If you don't see your instance after a few minutes, see Where Are My Instances? in the AWS Systems Manager User Guide.

6. In the left navigation pane, choose Automation.

7. Choose Execute automation.

8. Choose the button beside the AWSEC2-CloneInstanceAndUpgradeSQLServer SSM document, and then choose Next.

9. Ensure that the Simple execution option is selected.

10. Enter the requested parameters based on the following guidance.

   - InstanceId
     Type: String
     (Required) The instance running SQL Server 2008 R2 (or later).
   - IamInstanceProfile
     Type: String
     (Required) The IAM instance profile.
   - SnapshotId
     Type: String
     (Required) The Snapshot ID for SQL Server 2016 installation media.
   - SubnetId
     Type: String
     (Required) This is the subnet for the upgrade process and where your source EC2 instance resides. Verify that the subnet has outbound connectivity to AWS services, including Amazon S3, and also to Microsoft (in order to download patches).
   - KeepPreUpgradedBackUp
     Type: String
     (Optional) If this parameter is set to true, the automation retains the image created from the instance. The default setting is false.
   - RebootInstanceBeforeTakingImage
     Type: String
     (Optional) The default is false (no reboot). If this parameter is set to true, Systems Manager reboots the instance before creating an AMI for the upgrade.

11. After you have entered the parameters, choose Execute. When the automation begins, you can monitor the execution progress.

12. When Execution status shows Success, expand Outputs to view the AMI information. You can use the AMI ID to launch your SQL Server 2016 instance for the VPC of your choice.
13. Open the EC2 console. In the left navigation pane, choose AMIs. You should see the new AMI.
14. To verify that SQL Server 2016 has been successfully installed, choose the new AMI and choose Launch.
15. Choose the type of instance that you want for the AMI, the VPC and subnet that you want to deploy to, and the storage that you want to use. Because you’re launching the new instance from an AMI, the volumes are presented to you as an option to include within the new EC2 instance you are launching. You can remove any of these volumes, or you can add volumes.
16. Add a tag to help you identify your instance.
17. Add the security group or groups to the instance.
18. Choose Launch Instance.
19. Choose the tag name for the instance and select Connect under the Actions dropdown.
20. Verify that SQL Server 2016 is the new database engine on the new instance.

Migrating to Latest Generation Instance Types

The AWS Windows AMIs are configured with the default settings used by the Microsoft installation media with some customizations, including drivers and configurations, that support the latest generation instance types. However, if you are launching instances from custom Windows AMIs or from Windows AMIs provided by Amazon created before August 2018, we recommend that you follow the steps in this topic when migrating to the latest generation of EC2 instances, Nitro instances, including bare metal instances.

For more information, see Amazon EC2 Update — Additional Instance Types, Nitro System, and CPU Options.

Contents

- Part 1: Installing and Upgrading AWS PV Drivers (p. 507)
- Part 2: Installing and Upgrading ENA (p. 507)
- Part 3: Upgrading AWS NVMe Drivers (p. 508)
- Part 4: Updating EC2Config and EC2Launch (p. 508)
- Part 5: Installing the Serial Port Driver for Bare Metal Instances (p. 509)
- Part 6: Updating Power Management Settings (p. 509)
- Part 7: Updating Intel Chipset Drivers for New Instance Types (p. 509)
- (Alternative) Upgrading the AWS PV, ENA, and NVMe Drivers Using AWS Systems Manager (p. 510)

Note

Alternatively, you can use the AWSSupport-UpgradeWindowsAWSDrivers automation document to automate the procedures described in Part 1, Part 2, and Part 3. If you choose to use the automated procedure, see (Alternative) Upgrading the AWS PV, ENA, and NVMe Drivers Using AWS Systems Manager (p. 510), and then continue with Part 4 and Part 5.

Before you begin

This procedure assumes that you are currently running on a previous generation Xen-based instance type, such as an M4 or C4, and you are migrating to a latest generation instance type, such as an M5 or C5.

Note

When migrating to the latest generation instances, the static IP or custom DNS network settings on the existing ENI may be lost as the instance will default to a new Enhanced Networking Adapter device.
Before following the steps in this procedure, we recommend that you create a backup of the instance. From the EC2 console, choose the instance that requires the migration, open the context (right-click) menu, and choose **Instance State, Stop**.

**Warning**
When you stop an instance, the data on any instance store volumes is erased. To preserve data on instance store volumes, ensure that you back up the data to persistent storage.

Open the context (right-click) menu for the instance in the EC2 console, choose **Image**, and then choose **Create Image**.

**Note**
Parts 4 and 5 of these instructions can be completed after you migrate or change the instance type to the latest generation, such as M5 or C5. However, we recommend that you complete them before you migrate if you are migrating specifically to an EC2 Bare Metal instance type.

**Part 1: Installing and Upgrading AWS PV Drivers**

Though AWS PV drivers are not used in the Nitro system, you should still upgrade them if you are on previous versions of either Citrix PV or AWS PV. The latest AWS PV drivers resolve bugs in previous versions of the drivers that may appear while you are on a Nitro system, or if you need to migrate back to a Xen-based instance. As a best practice, we recommend always updating to the latest drivers for Windows instances on AWS.

Use the following procedure to perform an in-place upgrade of AWS PV drivers, or to upgrade from Citrix PV drivers to AWS PV drivers on Windows Server 2008 R2, Windows Server 2012, Windows Server 2012 R2, Windows Server 2016, or Windows Server 2019. For more information, see Upgrading PV Drivers on Your Windows Instances (p. 436).

**To perform an upgrade of or to AWS PV drivers**

1. Connect to the instance using Remote Desktop and prepare the instance for upgrade. Take all non-system disks offline before you perform the upgrade. If you are performing an in-place update of AWS PV drivers, this step is not required. Set non-essential services to **Manual** start-up in the Services console.
2. Download the latest driver package to the instance.
3. Extract the contents of the folder and run `AWSPVDriverSetup.msi`.

After running the MSI, the instance automatically reboots and upgrades the driver. The instance may not be available for up to 15 minutes.

After the upgrade is complete and the instance passes both health checks in the Amazon EC2 console, connect to the instance using Remote Desktop and verify that the new driver was installed. In Device Manager, under **Storage Controllers**, locate **AWS PV Storage Host Adapter**. Verify that the driver version is the same as the latest version listed in the Driver Version History table. For more information, see AWS PV Driver Package History (p. 432).

**Part 2: Installing and Upgrading ENA**

Upgrade to the latest Elastic Network Adapter driver to ensure that all network features are supported. If you launched your instance and it does not have enhanced networking already enabled, you must download and install the required network adapter driver on your instance. Then, set the enaSupport instance attribute to **activate enhanced networking**. You can only enable this attribute on supported instance types and only if the ENA driver is installed. For more information, see Enabling Enhanced Networking with the Elastic Network Adapter (ENA) on Windows Instances (p. 701).

1. Download the latest driver to the instance.
2. Extract the zip archive.
3. Install the driver by running the install.ps1 PowerShell script from the extracted folder.
   
   **Note**
   To avoid installation errors, run the install.ps1 script as an administrator.
4. Check if your AMI has enaSupport activated. If not, continue by following the documentation at Enabling Enhanced Networking with the Elastic Network Adapter (ENA) on Windows Instances (p. 701).

**Part 3: Upgrading AWS NVMe Drivers**

AWS NVMe drivers are used to interact with Amazon EBS and SSD instance store volumes that are exposed as NVMe block devices in the Nitro system for better performance.

**Important**

The following instructions are modified specifically for when you install or upgrade AWS NVMe on a previous generation instance with the intention to migrate the instance to the latest generation instance type.

1. Download the latest driver package to the instance.
2. Extract the zip archive.
3. Install the driver by running dpinst.exe.
4. Open a PowerShell session and run the following command:

   ```
   start rundll32.exe sppnp.dll,Sysprep_Generalize_Pnp -wait
   ```
   
   **Note**
   This command only runs sysprep on the driver devices. It does not run the full sysprep preparation.
5. For Windows Server 2008 R2 and Windows Server 2012, shut down the instance, change the instance type to a latest generation instance and start it, then proceed to Part 4. If you start the instance again on a previous generation instance type before migrating to a latest generation instance type, it will not boot. For other supported Windows AMIs, you can change the instance type anytime after the device sysprep.

**Part 4: Updating EC2Config and EC2Launch**

For Windows instances, the latest EC2Config and EC2Launch utilities provide additional functionality and information when running on the Nitro system, including on EC2 Bare Metal. By default, the EC2Config service is included in AMIs prior to Windows Server 2016. EC2Launch replaces EC2Config on Windows Server 2016 and later AMIs.

When the EC2Config and EC2Launch services are updated, new Windows AMIs from AWS include the latest version of the service. However, you must update your own Windows AMIs and instances with the latest version of EC2Config and EC2Launch.

**To install or update EC2Config**

1. Download and unzip the EC2Config Installer.
2. Run EC2Install.exe. For a complete list of options, run EC2Install with the /? option. By default, setup displays prompts. To run the command with no prompts, use the /quiet option.

   For more information, see Installing the Latest Version of EC2Config (p. 406).
**To install or update EC2Launch**

1. If you have already installed and configured EC2Launch on an instance, make a backup of the EC2Launch configuration file. The installation process does not preserve changes in this file. By default, the file is located in the `C:\ProgramData\Amazon\EC2-Windows\Launch\Config` directory.
2. Download `EC2-Windows-Launch.zip` to a directory on the instance.
3. Download `install.ps1` to the same directory where you downloaded `EC2-Windows-Launch.zip`.
4. Run `install.ps1`.

   **Note**
   
   To avoid installation errors, run the `install.ps1` script as an administrator.

5. If you made a backup of the EC2Launch configuration file, copy it to the `C:\ProgramData\Amazon\EC2-Windows\Launch\Config` directory.

For more information, see Configuring a Windows Instance Using EC2Launch (p. 395).

**Part 5: Installing the Serial Port Driver for Bare Metal Instances**

The `i3.metal` instance type uses a PCI-based serial device rather than an I/O port-based serial device. The latest Windows AMIs automatically use the PCI-based serial device and have the serial port driver installed. If you are not using an instance launched from an Amazon-provided Windows AMI dated 2018.04.11 or later, you must install the Serial Port Driver to enable the serial device for EC2 features such as Password Generation and Console Output. The latest EC2Config and EC2Launch utilities also support `i3.metal` and provide additional functionality. Follow the steps in Part 4, if you have not yet done so.

**To install the serial port driver**

1. Download the serial driver package to the instance.
2. Extract the contents of the folder, open the context (right-click) menu for `aws_ser.INF`, and choose `install`.
3. Choose `Okay`.

**Part 6: Updating Power Management Settings**

The following update to power management settings sets displays to never turn off, which allows for graceful OS shutdowns on the Nitro system. All Windows AMIs provided by Amazon as of 2018.11.28 already have this default configuration.

1. Open a command prompt or PowerShell session.
2. Run the following commands:

   ```
   powercfg /setacvalueindex 381b4222-f694-41f0-9685-ff5bb260df2e 7516b95f-f776-4464-8c53-06167f40cc99 3c0bc021-c8a8-4e07-a973-6b14cbb2b7e 0
   powercfg /setacvalueindex 8c5e7fda-e8bf-4a96-9a85-a6e23a8c635c 7516b95f-f776-4464-8c53-06167f40cc99 3c0bc021-c8a8-4e07-a973-6b14cbb2b7e 0
   powercfg /setacvalueindex a1841308-3541-4fab-bc81-f71556f20b4a 7516b95f-f776-4464-8c53-06167f40cc99 3c0bc021-c8a8-4e07-a973-6b14cbb2b7e 0
   ```

**Part 7: Updating Intel Chipset Drivers for New Instance Types**

The `u-6tb1.metal`, `u-9tb1.metal`, and `u-12tb1.metal` instance types use hardware that requires chipset drivers that were not previously installed on Windows AMIs. If you are not using an instance
launched from an Amazon-provided Windows AMI dated 2018.11.19 or later, you must install the drivers using the Intel Chipset INF Utility.

To install the chipset drivers
1. Download the chipset utility to the instance.
2. Extract the files.
3. Run SetupChipset.exe.
4. Accept the Intel software license agreement and install the chipset drivers.
5. Reboot the instance.

(Alternative) Upgrading the AWS PV, ENA, and NVMe Drivers Using AWS Systems Manager

The AWSSupport-UpgradeWindowsAWSDrivers automation document automates the steps described in Part 1, Part 2, and Part 3. This method can also repair an instance where the driver upgrades have failed.

The AWSSupport-UpgradeWindowsAWSDrivers automation document upgrades or repairs storage and network AWS drivers on the specified EC2 instance. The document attempts to install the latest versions of AWS drivers online by calling the AWS Systems Manager Agent (SSM Agent). If SSM Agent is not contactable, the document can perform an offline installation of the AWS drivers if explicitly requested.

Note
This procedure will fail on a domain controller. To update drivers on a domain controller, see Upgrade a Domain Controller (AWS PV Upgrade) (p. 437).

To automatically upgrade the AWS PV ENA, and NVMe drivers using AWS Systems Manager
2. Choose Automation, Execute Automation.
3. Choose the AWSSupport-UpgradeWindowsAWSDrivers automation document and then configure the following options in the Input Parameters section:

   Instance ID
   Enter the unique ID of the instance to upgrade.

   AllowOffline
   (Optional) Choose one of the following options:
   • True — Choose this option to perform an offline installation. The instance is stopped and restarted during the upgrade process.
     Warning
     When you stop an instance, the data on any instance store volumes is erased. To preserve data on instance store volumes, ensure that you back up the data to persistent storage.
   • False — (Default) To perform an online installation, leave this option selected. The instance is restarted during the upgrade process.

   Important
   Online and offline upgrades create an AMI before attempting the upgrade operations. The AMI persists after the automation completes. Secure your access to the AMI, or delete it if it is no longer needed.
SubnetId

(Optional) Enter one of the following values:

- **SelectedInstanceSubnet** — (Default) The upgrade process launches the helper instance into the same subnet as the instance that is to be upgraded. The subnet must allow communication to the Systems Manager endpoints (`ssm.*`).
- **CreateNewVPC** — The upgrade process launches the helper instance into a new VPC. Use this option if you’re not sure whether the target instance’s subnet allows communication to the `ssm.*` endpoints. Your IAM user must have permission to create a VPC.
- **A specific subnet ID** — Specify the ID of a specific subnet into which to launch the helper instance. The subnet must be in the same Availability Zone as the instance that is to be upgraded, and it must allow communication with the `ssm.*` endpoints.

4. Choose **Execute automation**.
5. Allow the upgrade to complete. It could take up to 10 minutes to complete an online upgrade, and up to 25 minutes to complete an offline upgrade.

### Windows to Linux Replatforming Assistant for Microsoft SQL Server Databases

The Windows to Linux replatforming assistant for Microsoft SQL Server Databases service is a scripting tool. It helps you move existing Microsoft SQL Server workloads from a Windows to a Linux operating system. You can use the replatforming assistant with any Windows Server virtual machines (VMs) hosted in the cloud, or with on-premises environments running Microsoft SQL Server 2008 and later. The tool checks for common incompatibilities, exports databases from the Windows VM, and imports into an EC2 instance running Microsoft SQL Server 2017 on Ubuntu 16.04. The automated process results in a ready-to-use Linux VM configured with your selected SQL Server databases that can be used for experimenting and testing.

**Contents**
- Concepts (p. 511)
- Related Services (p. 512)
- How Windows to Linux Replatforming Assistant for Microsoft SQL Server Works (p. 512)
- Components (p. 512)
- Setting Up (p. 512)
- Getting Started (p. 514)

### Concepts

The following terminology and concepts are central to your understanding and use of the Windows to Linux replatforming assistant for Microsoft SQL Server Databases.

**Backup**

A Microsoft SQL Server backup copies data or log records from a Microsoft SQL Server database or its transaction log to a backup device, such as a disk. For more information, see [Backup Overview (Microsoft SQL Server)](https://docs.microsoft.com/en-us/sql/)

**Restore**

A logical and meaningful sequence for restoring a set of Microsoft SQL Server backups. For more information, see [Restore and Recovery Overview (Microsoft SQL Server)](https://docs.microsoft.com/en-us/sql/)

**Replatform**
A Microsoft SQL Server database can be replatformed from an EC2 Windows instance to an EC2 Linux instance running Microsoft SQL Server. It can also be replatformed to the VMware Cloud running Microsoft SQL Server Linux on AWS.

**Related Services**

AWS Systems Manager (Systems Manager) gives you visibility and control of your infrastructure on AWS. The Windows to Linux replatforming assistant for Microsoft SQL Server Databases uses Systems Manager to move your Microsoft SQL databases to Microsoft SQL Server on EC2 Linux. For more information about Systems Manager, see the AWS Systems Manager User Guide.

**How Windows to Linux Replatforming Assistant for Microsoft SQL Server Works**

Windows to Linux replatforming assistant for Microsoft SQL Server Databases allows you to migrate your Microsoft SQL Server databases from an on-premises environment or from an EC2 Windows instance to Microsoft SQL Server 2017 on EC2 Linux using backup and restore. For the destination EC2 Linux instance, you provide either the EC2 instance ID or the EC2 instance type with the subnet ID and EC2 Key Pair.

When you execute the PowerShell script for the Windows to Linux replatforming assistant for Microsoft SQL Server Databases on the source Microsoft SQL Server databases, the Windows instance backs up the databases to an encrypted Amazon Simple Storage Service (S3) storage bucket. It then restores the backups to an existing Microsoft SQL Server on EC2 Linux instance, or it launches a new Microsoft SQL Server on EC2 Linux instance and restores the backups to the newly created instance. This process can be used to replatform your 2-tier databases running enterprise applications. It also enables you to replicate your database to Microsoft SQL Server on Linux to test the application while the source Microsoft SQL Server remains online. After testing, you can schedule application downtime and rerun the PowerShell backup script during your final cutover.

The entire replatforming process can also be automated and run unattended. You can run the Systems Manager SSM document AWSEC2-SQLServerDBRestore to import your existing database backup files into Microsoft SQL Server on EC2 Linux without using the PowerShell backup script.

**Components**

The Windows to Linux replatforming assistant for Microsoft SQL Server Databases script consists of two main components:

1. An AWS-signed PowerShell backup script, which backs up on-premises Microsoft SQL Server databases to an Amazon S3 storage bucket. It then invokes the SSM Automation document AWSEC2-SQLServerDBRestore to restore the backups to a Microsoft SQL Server on EC2 Linux instance.
2. An SSM Automation document named AWSEC2-SQLServerDBRestore, which restores database backups to Microsoft SQL Server on EC2 Linux. This automation restores Microsoft SQL Server database backups stored in Amazon S3 to Microsoft SQL Server 2017 running on an EC2 Linux instance. You can provide your own EC2 instance running Microsoft SQL Server 2017 Linux, or the automation launches and configures a new EC2 instance with Microsoft SQL Server 2017 on Ubuntu 16.04. The automation supports the restoration of full, differential, and transactional log backups, and accepts multiple database backup files. The automation automatically restores the most recent valid backup of each database in the files provided. For more information, see AWSEC2-SQLServerDBRestore.

**Setting Up**

This section covers the steps necessary to run the Windows to Linux replatforming script.
Prerequisites

In order to run the Windows to Linux replatforming assistant for Microsoft SQL Server Databases script, you must do the following:

1. **Install the AWS PowerShell Module**

   To install the AWS PowerShell module, follow the steps listed in Setting up the AWS Tools for PowerShell on a Windows-Based Computer. We recommend that you use PowerShell 3.0 or later for the backup script to work properly.

2. **Install the Windows to Linux Replatforming Assistant PowerShell Backup Script**

   In order to run the Windows to Linux replatforming assistant, download the PowerShell backup script: `MigrateSQLServerToEC2Linux.ps1`.

3. **Add an AWS User Profile to the AWS SDK Store**

   To add and configure the AWS user profile, see the steps listed in Managing Profiles in the AWS Tools for PowerShell User Guide. Set the following IAM policy for your user profile. You can also add these permissions as an inline policy under your AWS user account using the IAM console.

   ```json
   {
   "Version": "2012-10-17",
   "Statement": [
   {
   "Effect": "Allow",
   "Action": [
   "ec2:RebootInstances",
   "ec2:DescribeInstanceStatus",
   "ec2:DescribeInstances",
   "ec2:CreateTags",
   "ec2:RunInstances",
   "ec2:DescribeImages",
   "iam:PassRole",
   "ssm:StartAutomationExecution",
   "ssm:DescribeInstanceInformation",
   "ssm:ListCommandInvocations",
   "ssm:ListCommands",
   "ssm:SendCommand",
   "ssm:GetAutomationExecution",
   "ssm:GetCommandInvocation",
   "s3:PutEncryptionConfiguration",
   "s3:CreateBucket",
   "s3:ListBucket",
   "s3:PutObject",
   "s3:GetObject",
   "s3:DeleteObject",
   "s3:DeleteBucket"
   ],
   "Resource": "*"
   }
   ]
   }
   ```
4. Create an IAM Instance Profile Role

To create an IAM instance profile role in order to run Systems Manager on EC2 Linux, see the steps listed under Create an Instance Profile for Systems Manager in the AWS Systems Manager User Guide.

Prerequisites for Replatforming to an Existing EC2 Instance

To replatform to an existing instance running Microsoft SQL Server 2017 on Linux, you must:

1. Configure the EC2 instance with an AWS Identity and Access Management (IAM) instance profile and attach the AmazonSSMManagedInstanceCore managed policy.

   For information about creating an IAM instance profile for Systems Manager and attaching it to an instance, see the following topics in the AWS Systems Manager User Guide:
   - Create an Instance Profile for Systems Manager
   - Attach an IAM Instance Profile to an Amazon EC2 Instance

2. Verify that SSM Agent is installed on your EC2 instance. For more information, see Installing and Configuring SSM Agent on Windows Instances in the AWS Systems Manager User Guide.

3. Verify that the EC2 instance has enough free disk space to download and restore the Microsoft SQL Server backups.

Getting Started

This section contains the PowerShell parameter definitions and scripts for replatforming your databases. For more information about how to use PowerShell scripts, see PowerShell.

Topics
- Running the Windows to Linux Replatforming Assistant for Microsoft SQL Server Script (p. 514)
- Parameters (p. 515)

Running the Windows to Linux Replatforming Assistant for Microsoft SQL Server Script

The following common scenarios and example PowerShell scripts demonstrate how to replatform your Microsoft SQL Server databases using Windows to Linux replatforming assistant for Microsoft SQL Server Databases.

Important

The Windows to Linux Replatforming Assistant for Microsoft SQL Server Databases resets the SQL Server server administrator (SA) user password on the target instance every time that it is run. After the replatform process is complete, you must set your own SA user password before you can connect to the target SQL Server instance.

Syntax

The Windows to Linux replatforming assistant for Microsoft SQL Server Databases script adheres to the syntax shown in the following example.

```
PS C:\> C:\MigrateSQLServerToEC2Linux.ps1 [-SqlServerInstanceName] <String> [-DBNames] <Object[]> [-MigrateAllDBs] [PathForBackup] <String> [-SetSourceDBModeReadOnly] [-IamInstanceProfileName] <String> [-AWSRegion] <String> [-EC2InstanceId] <String> [-EC2InstanceType] <String> [-EC2KeyPair] <String>
```
Example 1: Move a database to an EC2 instance

The following example shows how to move a database named AdventureDB to an EC2 Microsoft SQL Server on Linux instance, with an instance ID of i-024689abcdef, from the Microsoft SQL Server Instance named MSSQLSERVER. The backup directory to be used is D:\Backup and the AWS Region is us-east-2.

```
PS C:\> ./MigrateSQLServerToEC2Linux.ps1 - SQLServerInstanceName MSSQLSERVER - EC2InstanceId i-024689abcdef -DBNames AdventureDB -PathForBackup D:\Backup -AWSRegion us-east-2 -IamInstanceProfileName AmazonSSMManagedInstanceCore
```

Example 2: Move a database to an EC2 instance using the AWS credentials profile

The following example shows how to move the database in Example 1 using the AWS credentials profile: DBMigration.

```
PS C:\> ./MigrateSQLServerToEC2Linux.ps1 - SQLServerInstanceName MSSQLSERVER - EC2InstanceId i-024689abcdef -DBNames AdventureDB -PathForBackup D:\Backup -AWSRegion us-east-2 -AWSProfileName DBMigration -IamInstanceProfileName AmazonSSMManagedInstanceCore
```

Example 3: Move a database to a new m5.large type instance

The following example shows how to create an m5.large type EC2 Linux instance in subnet-abc127 using the Key Pair customer-ec2-keypair and then moving AdventureDB and TestDB to the new instance from the database used in Examples 1 and 2.

```
PS C:\> ./MigrateSQLServerToEC2Linux.ps1 -EC2InstanceType m5.large -SubnetId subnet-abc127 -EC2KeyPair customer-ec2-keypair -DBNames AdventureDB,TestDB -PathForBackup D:\Backup -AWSRegion us-east-2 -AWSProfileName DBMigration -IamInstanceProfileName AmazonSSMManagedInstanceCore
```

Example 4: Move all databases to a new m5.large type instance

The following example shows how to create an m5.large type EC2 Linux instance in subnet-abc127 using the Key Pair customer-ec2-keypair and then migrating all databases to the instance from databases used in Examples 1 and 2.

```
PS C:\> ./MigrateSQLServerToEC2Linux.ps1 -EC2InstanceType m5.large -SubnetId subnet-abc127 -EC2KeyPair customer-ec2-keypair -MigrateAllDBs -PathForBackup D:\Backup -AWSRegion us-east-2 -AWSProfileName DBMigration -IamInstanceProfileName AmazonSSMManagedInstanceCore
```

Parameters

The following parameters are used by the PowerShell script to replatform your Microsoft SQL Server databases.

- `SqlServerInstanceName`

---

515
The name of the Microsoft SQL Server instance to be backed up. If a value for SqlServerInstanceName is not provided, $env:ComputerName is used by default.

Type: String
Required: No

-DBNames

The names of the databases to be backed up and restored. Specify the names of the databases in a comma-separated list (for example, adventureDB, universityDB). Either the DBNames or MigrateAllDBs parameter is required.

Type: Object
Required: No

-MigrateAllDBs

This switch is disabled by default. If this switch is enabled, the automation migrates all databases except for the system databases (master, msdb, tempdb). Either the DBNames or MigrateAllDBs parameter is required.

Type: SwitchParameter
Required: No

-PathForBackup

The path where the full backup is stored.

Type: String
Required: Yes

-SetSourceDBModeReadOnly

This switch is disabled by default. If this switch is enabled, it makes the database read-only during migration.

Type: SwitchParameter
Required: No

-IamInstanceProfileName

Enter the AWS IAM instance role with permissions to run Systems Manager Automation on your behalf. See Getting Started with Automation in the AWS Systems Manager User Guide.

Type: String
Required: Yes

-AWSRegion

Enter the AWS Region where your Amazon S3 buckets are created to store database backups.

Type: String
Required: Yes
-EC2InstanceId

To restore Microsoft SQL Server databases to an existing EC2 instance running Microsoft SQL Server Linux, enter the instance ID of the instance. Make sure that the EC2 instance already has the AWS Systems Manager SSM Agent installed and running.

Type: String
Required: No

-EC2InstanceType

To restore Microsoft SQL Server databases to a new EC2 Linux instance, enter the instance type of the instance to be launched.

Type: String
Required: No

-EC2KeyPair

To restore Microsoft SQL Server databases to a new EC2 Linux instance, enter the name of the EC2 Key Pair to be used to access the instance. This parameter is recommended if you are creating a new EC2 Linux instance.

Type: String
Required: No

-SubnetId

This parameter is required when creating a new EC2 Linux instance. When creating a new EC2 Linux instance, if SubnetId is not provided, the AWS user default subnet is used to launch the EC2 Linux instance.

Type: String
Required: No

-AWSProfileName

The name of the AWS profile that the automation uses when connecting to AWS services. For more information on the required IAM user permissions, see Getting Started with Automation in the AWS Systems Manager User Guide. If a profile is not entered, the automation uses your default AWS profile.

Type: String
Required: No

-AWSProfileLocation

The location of the AWS Profile if the AWS Profile is not stored in the default location.

Type: String
Required: No

-GeneratePresignedUrls

This parameter is only used when replatforming to non-EC2 instances, such as to VMware Cloud on AWS or on-premises VMs.
Troubleshooting an Upgrade

AWS provides upgrade support for issues or problems with the Upgrade Helper Service, an AWS utility that helps you perform in-place upgrades involving Citrix PV drivers.

After the upgrade, the instance might temporarily experience higher than average CPU utilization while the .NET Runtime Optimization service optimizes the .NET framework. This is expected behavior.

If the instance has not passed both status checks after several hours, check the following:

- If you upgraded to Windows Server 2008 and both status checks fail after several hours, the upgrade may have failed and be presenting a prompt to **Click OK** to confirm rolling back. Because the console is not accessible at this state, there is no way to click the button. To get around this, perform a reboot via the Amazon EC2 console or API. The reboot takes ten minutes or more to initiate. The instance might become available after 25 minutes.
- Remove applications or server roles from the server and try again.

If the instance does not pass both status checks after removing applications or server roles from the server, do the following:

- Stop the instance and attach the root volume to another instance. For more information, see the description of how to stop and attach the root volume to another instance in "Waiting for the metadata service" (p. 1074).
- Analyze Windows Setup log files and event logs for failures.

For other issues or problems with an operating system upgrade or migration, we recommend reviewing the articles listed in Before You Begin an In-Place Upgrade (p. 496).

Identify EC2 Windows Instances

Your application might need to determine whether it is running on an EC2 instance.

For information about identifying Linux instances, see Identify EC2 Linux Instances in the Amazon EC2 User Guide for Linux Instances.

Inspecting the Instance Identity Document

For a definitive and cryptographically verified method of identifying an EC2 instance, check the instance identity document, including its signature. These documents are available on every EC2 instance at the local, non-routable address [http://169.254.169.254/latest/dynamic/instance-identity/](http://169.254.169.254/latest/dynamic/instance-identity/). For more information, see Instance Identity Documents (p. 487).
Inspecting the System UUID

You can get the system UUID and look for the presence of the characters "EC2" in the beginning octet of the UUID. This method to determine whether a system is an EC2 instance is quick but potentially inaccurate because there is a small chance that a system that is not an EC2 instance could have a UUID that starts with these characters. Furthermore, EC2 instances using SMBIOS 2.4 might represent the UUID in little-endian format, therefore the "EC2" characters do not appear at the beginning of the UUID.

Example : Get the UUID using WMI or Windows PowerShell

Use the Windows Management Instrumentation command line (WMIC) as follows:

```
wmic path win32_computersystemproduct get uuid
```

Alternatively, if you’re using Windows PowerShell, use the `Get-WmiObject` cmdlet as follows:

```
PS C:\> Get-WmiObject -query "select uuid from Win32_ComputerSystemProduct" | Select UUID
```

In the following example output, the UUID starts with "EC2", which indicates that the system is probably an EC2 instance.

```
EC2AE145-D1DC-13B2-94ED-01234ABCDEF
```

For instances using SMBIOS 2.4, the UUID might be represented in little-endian format; for example:

```
45E12ABC-DCD1-B213-94ED-01234ABCDEF
```
Amazon Elastic Graphics

Amazon Elastic Graphics provides flexible, low-cost, and high performance graphics acceleration for your Windows instances. Elastic Graphics accelerators come in multiple sizes and are a low-cost alternative to using GPU graphics instance types (such as G2 and G3). You have the flexibility to choose an instance type that meets the compute, memory, and storage needs of your application. Then, choose the accelerator for your instance that meets the graphics requirements of your workload.

Elastic Graphics is suited for applications that require a small or intermittent amount of additional graphics acceleration, and that use OpenGL graphics support. If you need access to full, directly attached GPUs and use of DirectX, CUDA, or Open Computing Language (OpenCL) parallel computing frameworks, use an accelerated computing instance type instance instead. For more information, see Windows Accelerated Computing Instances (p. 178).

Contents
- Elastic Graphics Basics (p. 520)
- Pricing for Elastic Graphics (p. 522)
- Elastic Graphics Limitations (p. 522)
- Working with Elastic Graphics (p. 522)
- Using CloudWatch Metrics to Monitor Elastic Graphics (p. 527)
- Troubleshooting (p. 529)

Elastic Graphics Basics

To use Elastic Graphics, launch a Windows instance and specify an accelerator type for the instance during launch. AWS finds available Elastic Graphics capacity and establishes a network connection between your instance and the Elastic Graphics accelerator.

**Note**
Bare metal instances are not supported.

The following instance types support Elastic Graphics accelerators:

- C3 | C4 | C5 | C5d | C5n
- D2
- H1
- I3 | I3en
- M3 | M4 | M5 | M5d
- P2 | P3 | P3dn
- R3 | R4 | R5 | R5d
- t2.medium or larger | t3.medium or larger
- X1 | X1e
- z1d

The following Elastic Graphics accelerators are available. You can attach any Elastic Graphics accelerator to any supported instance type.
Elastic Graphics Basics

<table>
<thead>
<tr>
<th>Elastic Graphics accelerator</th>
<th>Graphics memory (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eg1.medium</td>
<td>1</td>
</tr>
<tr>
<td>eg1.large</td>
<td>2</td>
</tr>
<tr>
<td>eg1.xlarge</td>
<td>4</td>
</tr>
<tr>
<td>eg1.2xlarge</td>
<td>8</td>
</tr>
</tbody>
</table>

An Elastic Graphics accelerator does not form part of the hardware of your instance. Instead, it is network-attached through a network interface, known as the Elastic Graphics network interface. When you launch an instance with graphics acceleration, the Elastic Graphics network interface is created in your VPC for you.

The Elastic Graphics network interface is created in the same subnet and VPC as your instance and is assigned a private IPv4 address from that subnet. The accelerator attached to your Amazon EC2 instance is allocated from a pool of available accelerators in the same Availability Zone as your instance.

Elastic Graphics accelerators support the API standards for OpenGL 4.3 API and earlier, which can be used for batch applications or 3D-graphics acceleration. An Amazon-optimized OpenGL library on your instance detects the attached accelerator. It directs OpenGL API calls from your instance to the accelerator, which then processes the requests and returns the results. Traffic between the instance and the accelerator uses the same bandwidth as the instance's network traffic so we recommend that you have adequate network bandwidth available. Consult your software vendor for any OpenGL compliance and version questions.

By default, the default security group for your VPC is associated with the Elastic Graphics network interface. The Elastic Graphics network traffic uses the TCP protocol and port 2007. Ensure that the security group for your instance allows for this. For more information, see Configuring Your Security Groups (p. 522).
Pricing for Elastic Graphics

You are charged for each second that an Elastic Graphics accelerator is attached to an instance in the running state when the accelerator is in the Ok state. You are not charged for an accelerator attached to an instance that is in the pending, stopping, stopped, shutting-down, or terminated state. You are also not charged when an accelerator is in the Unknown or Impaired state.

Pricing for accelerators is available at On-Demand rates only. You can attach an accelerator to a Reserved, Scheduled, or Spot Instance. The On-Demand price for the accelerator applies in all cases.

For more information, see Amazon Elastic Graphics Pricing.

Elastic Graphics Limitations

Before you start using Elastic Graphics accelerators, be aware of the following limitations:

- You can attach accelerators only to Windows instances with Microsoft Windows Server 2012 R2 or later. Linux instances are currently not supported.
- You can attach one accelerator to an instance at a time, and only during instance launch.
- You can’t share an accelerator between instances.
- You can’t detach an accelerator from an instance or transfer it to another instance. If you no longer require an accelerator, you must terminate your instance. To change the accelerator type, create an AMI from your instance, terminate the instance, and launch a new instance with a different accelerator specification.
- The only supported versions of the OpenGL API are 4.3 and earlier. DirectX, CUDA, and OpenCL are not supported.
- The Elastic Graphics accelerator is not visible or accessible through the device manager of your instance.
- You can’t reserve or schedule accelerator capacity.
- You can’t attach accelerators to instances in EC2-Classic.

Working with Elastic Graphics

You can launch an instance and associate it with an Elastic Graphics accelerator during launch. You must then manually install the necessary libraries on your instance that enable communication with the accelerator.

Tasks

- Configuring Your Security Groups (p. 522)
- Launching an Instance with an Elastic Graphics accelerator (p. 523)
- Installing the Required Software for Elastic Graphics (p. 524)
- Verifying Elastic Graphics Functionality on Your Instance (p. 524)
- Viewing Elastic Graphics Information (p. 526)
- Submitting Feedback (p. 527)

Configuring Your Security Groups

If you use the Amazon EC2 console to launch your instance with an Elastic Graphics accelerator and create a security group for you, the console adds the inbound and outbound rules that are required to
allow traffic on the Elastic Graphics port. If you are launching your instance using the AWS CLI or an SDK, you must ensure that your security group allows traffic on the Elastic Graphics port.

To create a security group for Elastic Graphics

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Security Groups, Create Security Group.
3. Provide a name for your security group, such as “Elastic Graphics security group”, and a description for the security group. Select the VPC that you will use to launch your instance with an Elastic Graphics accelerator.
4. Create an inbound security group rule as follows:
   a. On the Inbound tab, choose Add Rule
   b. For Type, choose Elastic Graphics. For Source, choose Custom and type the ID of the security group.
5. Create an outbound security group rule as follows:
   a. On the Outbound tab, choose Add Rule
   b. For Type, choose All TCP. For Destination, choose Custom and type the ID of the security group.
6. Choose Create.

For more information, see Amazon EC2 Security Groups for Windows Instances (p. 583).

Launching an Instance with an Elastic Graphics accelerator

You can associate an Elastic Graphics accelerator to an instance during launch. If the launch fails, the following are possible reasons:

- Insufficient Elastic Graphics accelerator capacity
- Exceeded limit on Elastic Graphics accelerators in the Region
- Not enough private IPv4 addresses in your VPC to create a network interface for the accelerator

For more information, see Elastic Graphics Limitations (p. 522).

To associate an Elastic Graphics accelerator during instance launch (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the dashboard, choose Launch Instance.
3. Select a Windows AMI and a supported instance type. For more information, see Elastic Graphics Basics (p. 520).
4. On the Configure Instance Details page, select a VPC and subnet in which to launch your instance.
6. (Optional) On the Add Storage and Add Tags pages, add volumes and tags as needed.
7. On the Configure Security Group page, you can let the console create a security group for you with the required inbound and outbound rules, or you can use the security group that you created manually in Configuring Your Security Groups (p. 522). Add additional security groups as needed.
8. Choose Review and Launch to review your instance options and then choose Launch.

To associate an Elastic Graphics accelerator during instance launch (AWS CLI)
You can use the `run-instances` AWS CLI command with the following parameter:

```
--elastic-gpu-specification Type=eg1.medium
```

For the `--security-group-ids` parameter, you must include a security group that has the required inbound and outbound rules. For more information, see Configuring Your Security Groups (p. 522).

To associate an Elastic Graphics accelerator during instance launch (Tools for Windows PowerShell)

Use the `New-EC2Instance` Tools for Windows PowerShell command.

### Installing the Required Software for Elastic Graphics

If you launched your instance using a current AWS Windows AMI, the required software is installed automatically during the first boot. If you launched your instance using Windows AMIs that do not automatically install the required software, you must install the required software on the instance manually.

**To install the required software for Elastic Graphics (if necessary)**

1. Connect to the instance.
2. Download the Elastic Graphics installer and open it. The installation manager connects to the Elastic Graphics endpoint and downloads the latest version of the required software.
3. Reboot the instance to verify.

### Verifying Elastic Graphics Functionality on Your Instance

The Elastic Graphics packages on your instance include tools that you can use to view the status of the accelerator, and to verify that OpenGL commands from your instance to the accelerator are functional.

If your instance was launched with an AMI that does not have the Elastic Graphics packages pre-installed, you can download and install them yourself. For more information, see Installing the Required Software for Elastic Graphics (p. 524).

**Contents**

- Using the Elastic Graphics Status Monitor (p. 524)
- Using the Elastic Graphics Command Line Tool (p. 525)

### Using the Elastic Graphics Status Monitor

You can use the status monitor tool to view information about the status of an attached Elastic Graphics accelerator. By default, this tool is available in the notification area of the taskbar in your Windows instance and shows the status of the graphics accelerator. The following are the possible values.

**Healthy**

The Elastic Graphics accelerator is enabled and healthy.

**Updating**

The status of the Elastic Graphics accelerator is currently updating. It might take a few minutes to display the status.
Out of service

The Elastic Graphics accelerator is out of service. To get more information about the error, choose Read More.

Using the Elastic Graphics Command Line Tool

You can use the Elastic Graphics command line tool, egcli.exe, to check the status of the accelerator. If there is a problem with the accelerator, the tool returns an error message.

To launch the tool, open a command prompt from within your instance and run the following command:

```
C:\Program Files\Amazon\EC2ElasticGPUs\manager\egcli.exe
```

The tool also supports the following parameters:

```
--json, -j
```

Indicates whether to show the JSON message. The possible values are true and false. The default is true.

```
--imds, -i
```

Indicates whether to check the instance metadata for the availability of the accelerator. The possible values are true and false. The default is true.

The following is example output. A status of OK indicates that the accelerator is enabled and healthy.

```
EG Infrastructure is available.
Instance ID egpu-f6d94dfa66df4883b284e96db7397ee6
Instance Type eg1.large
EG Version 1.0.0.885 (Manager) / 1.0.0.95 (OpenGL Library) / 1.0.0.69 (OpenGL Redirector)
EG Status: Healthy
JSON Message:
{
  "version": "2016-11-30",
  "status": "OK"
}
```

The following are the possible values for status:

OK

The Elastic Graphics accelerator is enabled and healthy.

UPDATING

The Elastic Graphics driver is being updated.

NEEDS_REBOOT

The Elastic Graphics driver has been updated and a reboot of the Amazon EC2 instance is required.

LOADING_DRIVER

The Elastic Graphics driver is being loaded.

CONNECTING_EGPU

The Elastic Graphics driver is verifying the connectivity with the Elastic Graphics accelerator.
Viewing Elastic Graphics Information

You can view information about the Elastic Graphics accelerator attached to your instance.

**To view information about an Elastic Graphics accelerator (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances** and select your instance.
3. On the **Description** tab, find **Elastic Graphics ID**. Choose the ID to view the following information about the Elastic Graphics accelerator:
   - Attachment State
   - Type
   - Health status

**To view information about an Elastic Graphics accelerator (AWS CLI)**

You can use the `describe-elastic-gpus` AWS CLI command:

```
aws ec2 describe-elastic-gpus
```

You can use the `describe-network-interfaces` AWS CLI command and filter by owner ID to view information about the Elastic Graphics network interface.

```
aws ec2 describe-network-interfaces --filters "Name=attachment.instance-owner-id,Values=amazon-elasticgpus"
```

**To view information about an Elastic Graphics accelerator (Tools for Windows PowerShell)**

Use the following commands:

- `Get-EC2ElasticGpu`
- `Get-EC2NetworkInterface`

**To view information about an Elastic Graphics accelerator using instance metadata**

1. Connect to your Windows instance that is using an Elastic Graphics accelerator.
2. Do one of the following:
   - From PowerShell, use the following cmdlet:
Submitting Feedback

You can submit feedback about your experience with Elastic Graphics so the team can make further improvements.

To submit feedback using the Elastic Graphics Status Monitor

1. In the notification area of the taskbar in your Windows instance, open the Elastic Graphics Status Monitor.
2. In the lower left corner, choose Feedback.
3. Enter your feedback and choose Submit.

Using CloudWatch Metrics to Monitor Elastic Graphics

You can monitor your Elastic Graphics accelerator using Amazon CloudWatch, which collects metrics about your accelerator performance. These statistics are recorded for a period of two weeks, so that you can access historical information and gain a better perspective on how your service is performing.

By default, Elastic Graphics accelerators send metric data to CloudWatch in 5-minute periods.

For more information about Amazon CloudWatch, see the Amazon CloudWatch User Guide.

Elastic Graphics Metrics

The AWS/ElasticGPUs namespace includes the following metrics for Elastic Graphics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPUConnectivityCheckFailed</td>
<td>Reports whether connectivity to the Elastic Graphics accelerator is active or has failed. A value of zero (0) indicates that the connection is active. A value of one (1) indicates a connectivity failure.</td>
</tr>
<tr>
<td></td>
<td><strong>Units:</strong> Count</td>
</tr>
<tr>
<td>GPUHealthCheckFailed</td>
<td>Reports whether the Elastic Graphics accelerator has passed a status health check in the last minute. A value of zero (0) indicates that the status check passed. A value of one (1) indicates a status check failure.</td>
</tr>
<tr>
<td></td>
<td><strong>Units:</strong> Count</td>
</tr>
</tbody>
</table>
### Elastic Graphics Dimensions

You can filter the metrics data for your Elastic Graphics accelerators using the following dimensions.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGPUId</td>
<td>Filters the data by the Elastic Graphics accelerator.</td>
</tr>
<tr>
<td>InstanceId</td>
<td>Filters the data by the instance to which the Elastic Graphics accelerator is attached.</td>
</tr>
</tbody>
</table>

### Viewing CloudWatch Metrics for Elastic Graphics

Metrics are grouped first by the service namespace, and then by the supported dimensions. You can use the following procedures to view the metrics for your Elastic Graphics accelerators.

**To view Elastic Graphics metrics using the CloudWatch console**

2. If necessary, change the Region. From the navigation bar, select the Region where your Elastic Graphics accelerator resides. For more information, see Regions and Endpoints.
3. In the navigation pane, choose **Metrics**.
4. For **All metrics**, select **Elastic Graphics, Elastic Graphics Metrics**.

**To view Elastic Graphics metrics (AWS CLI)**

Use the following `list-metrics` command:

```bash
aws cloudwatch list-metrics --namespace "AWS/ElasticGPUs"
```

### Creating CloudWatch Alarms to Monitor Elastic Graphics

You can create a CloudWatch alarm that sends an Amazon SNS message when the alarm changes state. An alarm watches a single metric over a time period you specify, and sends a notification to an Amazon SNS topic based on the value of the metric relative to a given threshold over a number of time periods.

For example, you can create an alarm that monitors the health of an Elastic Graphics accelerator and sends a notification when the graphics accelerator fails a status health check for three consecutive 5-minute periods.

**To create an alarm for an Elastic Graphics accelerator health status**

2. In the navigation pane, choose Alarms, Create Alarm.
4. Select the GPUHealthCheckFailed metric and choose Select metric.
5. Configure the alarm as follows:
   a. For Alarm details, type a name and description for your alarm. For Whenever, choose >= and type 1.
   b. For Actions, select an existing notification list or choose New list.
   c. Choose Create Alarm.

Troubleshooting

The following are common errors and troubleshooting steps.

Contents
- Investigating Application Performance Issues (p. 529)
  - OpenGL Rendering Performance Issues (p. 529)
  - Remote Access Performance Issues (p. 530)
- Resolving Unhealthy Status Issues (p. 531)
  - Stop and Start the Instance (p. 531)
  - Verify the Installed Components (p. 531)
  - Check the Elastic Graphics Logs (p. 531)

Investigating Application Performance Issues

Elastic Graphics uses the instance network to send OpenGL commands to a remotely attached graphics card. In addition, a desktop running an OpenGL application with an Elastic Graphics accelerator is usually accessed using a remote access technology. It is important to distinguish between a performance problem related to the OpenGL rendering or the desktop remote access technology.

OpenGL Rendering Performance Issues

The OpenGL rendering performance is determined by the number of OpenGL commands and frames generated on the remote instance.

Rendering performance may vary depending on the following factors:
- Elastic Graphics accelerator performance
- Network performance
- CPU performance
- Rendering model, scenario complexity
- OpenGL application behavior

An easy way to evaluate performance is to display the number of rendered frames on the remote instance. Elastic Graphics accelerators display a maximum of 25 FPS on the remote instance to achieve the best perceived quality while reducing network usage.

To show the number of frames produced
1. Open the following file in a text editor. If the file does not exist, create it.
2. Identify the [Application] section, or add it if it is not present, and add the following configuration parameter:

```
[Application]
show_fps=1
```

3. Restart the application and check the FPS again.

If the FPS reaches 15-25 FPS when updating the rendered scene, then the Elastic Graphics accelerator is performing at peak. Any other performance problems you experience are likely related to the remote access to the instance desktop. If that is the case, see the Remote Access Performance Issues section.

If the FPS number is lower than 15, you can try the following:

- Improve Elastic Graphics accelerator performance by selecting a more powerful graphics accelerator type.
- Improve overall network performance by using these tips:
  - Check the amount of incoming and outgoing bandwidth to and from the Elastic Graphics accelerator endpoint. The Elastic Graphics accelerator endpoint can be retrieved with the following PowerShell command:

  ```
  ```
  - The network traffic from the instance to the Elastic Graphics accelerator endpoint relates to the volume of commands the OpenGL application is producing.
  - The network traffic from the Elastic Graphics accelerator endpoint to the instance relates to the number of frames generated by the graphics accelerator.
  - If you see the network usage reaching the instances maximum network throughput, try using an instance with a higher network throughput allowance.
- Improve CPU performance:
  - Applications may require a lot of CPU resources in addition to what the Elastic Graphics accelerator requires. If Windows Task Manager is reporting a high usage of CPU resources, try using an instance with more CPU power.

### Remote Access Performance Issues

An instance with an attached Elastic Graphics accelerator can be accessed using different remote access technologies. Performance and quality may vary depending on:

- The remote access technology
- Instance performance
- Client performance
- Network latency and bandwidth between the client and the instance

Possible choices for the remote access protocol include:

- Microsoft Remote Desktop Connection
- NICE DCV
- VNC
For more information about optimization, see the specific protocol.

Resolving Unhealthy Status Issues

If the Elastic Graphics accelerator is in an unhealthy state, use the following troubleshooting steps to resolve the issue.

Stop and Start the Instance

If your Elastic Graphics accelerator is in an unhealthy state, stopping the instance and starting it again is the simplest option. For more information, see Stopping and Starting Your Instances (p. 384).

Warning
When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.

Verify the Installed Components

Open the Windows Control Panel and confirm that the following components are installed:

- Amazon Elastic Graphics Manager
- Amazon Elastic Graphics OpenGL Library
- Amazon EC2 Elastic GPUs OpenGL Redirector

If any of these items are missing, you must install them manually. For more information, see Installing the Required Software for Elastic Graphics (p. 524).

Check the Elastic Graphics Logs

Open the Windows Event Viewer, expand the Application and Services Logs section, and search for errors in the following event logs:

- EC2ElasticGPUs
- EC2ElasticGPUs GUI
Monitoring Amazon EC2

Monitoring is an important part of maintaining the reliability, availability, and performance of your Amazon Elastic Compute Cloud (Amazon EC2) instances and your AWS solutions. You should collect monitoring data from all of the parts in your AWS solutions so that you can more easily debug a multi-point failure if one occurs. Before you start monitoring Amazon EC2, however, you should create a monitoring plan that should include:

- What are your goals for monitoring?
- What resources will you monitor?
- How often will you monitor these resources?
- What monitoring tools will you use?
- Who will perform the monitoring tasks?
- Who should be notified when something goes wrong?

After you have defined your monitoring goals and have created your monitoring plan, the next step is to establish a baseline for normal Amazon EC2 performance in your environment. You should measure Amazon EC2 performance at various times and under different load conditions. As you monitor Amazon EC2, you should store a history of monitoring data that you've collected. You can compare current Amazon EC2 performance to this historical data to help you to identify normal performance patterns and performance anomalies, and devise methods to address them. For example, you can monitor CPU utilization, disk I/O, and network utilization for your EC2 instances. When performance falls outside your established baseline, you might need to reconfigure or optimize the instance to reduce CPU utilization, improve disk I/O, or reduce network traffic.

To establish a baseline you should, at a minimum, monitor the following items:

<table>
<thead>
<tr>
<th>Item to Monitor</th>
<th>Amazon EC2 Metric</th>
<th>Monitoring Agent/CloudWatch Logs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU utilization</td>
<td>CPUUtilization (p. 548)</td>
<td></td>
</tr>
<tr>
<td>Network utilization</td>
<td>NetworkIn (p. 548)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NetworkOut (p. 548)</td>
<td></td>
</tr>
<tr>
<td>Disk performance</td>
<td>DiskReadOps (p. 548)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DiskWriteOps (p. 548)</td>
<td></td>
</tr>
<tr>
<td>Disk Reads/Writes</td>
<td>DiskReadBytes (p. 548)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DiskWriteBytes (p. 548)</td>
<td></td>
</tr>
<tr>
<td>Memory utilization, disk swap utilization, disk space utilization, page file utilization, log collection</td>
<td>[Linux and Windows Server instances] Collect Metrics and Logs from Amazon EC2 Instances and On-Premises Servers with the CloudWatch Agent</td>
<td>[Migration from previous CloudWatch Logs agent on]</td>
</tr>
</tbody>
</table>
Automated and Manual Monitoring

AWS provides various tools that you can use to monitor Amazon EC2. You can configure some of these tools to do the monitoring for you, while some of the tools require manual intervention.

**Topics**
- Automated Monitoring Tools (p. 533)
- Manual Monitoring Tools (p. 534)

### Automated Monitoring Tools

You can use the following automated monitoring tools to watch Amazon EC2 and report back to you when something is wrong:

- **System Status Checks** - monitor the AWS systems required to use your instance to ensure they are working properly. These checks detect problems with your instance that require AWS involvement to repair. When a system status check fails, you can choose to wait for AWS to fix the issue or you can resolve it yourself (for example, by stopping and restarting or terminating and replacing an instance). Examples of problems that cause system status checks to fail include:
  - Loss of network connectivity
  - Loss of system power
  - Software issues on the physical host
  - Hardware issues on the physical host that impact network reachability

For more information, see [Status Checks for Your Instances (p. 535)](#).

- **Instance Status Checks** - monitor the software and network configuration of your individual instance. These checks detect problems that require your involvement to repair. When an instance status check fails, typically you will need to address the problem yourself (for example, by rebooting the instance or by making modifications in your operating system). Examples of problems that may cause instance status checks to fail include:
  - Failed system status checks
  - Misconfigured networking or startup configuration
  - Exhausted memory
  - Corrupted file system
  - Incompatible kernel

For more information, see [Status Checks for Your Instances (p. 535)](#).

- **Amazon CloudWatch Alarms** - watch a single metric over a time period you specify, and perform one or more actions based on the value of the metric relative to a given threshold over a number of time periods. The action is a notification sent to an Amazon Simple Notification Service (Amazon SNS) topic or Amazon EC2 Auto Scaling policy. Alarms invoke actions for sustained state changes only. CloudWatch alarms will not invoke actions simply because they are in a particular state; the state...
must have changed and been maintained for a specified number of periods. For more information, see Monitoring Your Instances Using CloudWatch (p. 546).

- **Amazon CloudWatch Events** - automate your AWS services and respond automatically to system events. Events from AWS services are delivered to CloudWatch Events in near real time, and you can specify automated actions to take when an event matches a rule you write. For more information, see What is Amazon CloudWatch Events?

- **Amazon CloudWatch Logs** - monitor, store, and access your log files from Amazon EC2 instances, AWS CloudTrail, or other sources. For more information, see the Amazon CloudWatch Logs User Guide.

- **Amazon EC2 Monitoring Scripts** - Perl scripts that can monitor memory, disk, and swap file usage in your instances. For more information, see Monitoring Memory and Disk Metrics for Amazon EC2 Linux Instances.

- **AWS Management Pack for Microsoft System Center Operations Manager** - links Amazon EC2 instances and the Windows or Linux operating systems running inside them. The AWS Management Pack is an extension to Microsoft System Center Operations Manager. It uses a designated computer in your datacenter (called a watcher node) and the Amazon Web Services APIs to remotely discover and collect information about your AWS resources. For more information, see AWS Management Pack for Microsoft System Center (p. 992).

**Manual Monitoring Tools**

Another important part of monitoring Amazon EC2 involves manually monitoring those items that the monitoring scripts, status checks, and CloudWatch alarms don't cover. The Amazon EC2 and CloudWatch console dashboards provide an at-a-glance view of the state of your Amazon EC2 environment.

- **Amazon EC2 Dashboard** shows:
  - Service Health and Scheduled Events by Region
  - Instance state
  - Status checks
  - Alarm status
  - Instance metric details (In the navigation pane choose **Instances**, select an instance, and choose the **Monitoring** tab)
  - Volume metric details (In the navigation pane choose **Volumes**, select a volume, and choose the **Monitoring** tab)

- **Amazon CloudWatch Dashboard** shows:
  - Current alarms and status
  - Graphs of alarms and resources
  - Service health status

In addition, you can use CloudWatch to do the following:

- Graph Amazon EC2 monitoring data to troubleshoot issues and discover trends
- Search and browse all your AWS resource metrics
- Create and edit alarms to be notified of problems
- See at-a-glance overviews of your alarms and AWS resources

**Best Practices for Monitoring**

Use the following best practices for monitoring to help you with your Amazon EC2 monitoring tasks.

- Make monitoring a priority to head off small problems before they become big ones.
• Create and implement a monitoring plan that collects monitoring data from all of the parts in your AWS solution so that you can more easily debug a multi-point failure if one occurs. Your monitoring plan should address, at a minimum, the following questions:
  • What are your goals for monitoring?
  • What resources you will monitor?
  • How often you will monitor these resources?
  • What monitoring tools will you use?
  • Who will perform the monitoring tasks?
  • Who should be notified when something goes wrong?
• Automate monitoring tasks as much as possible.
• Check the log files on your EC2 instances.

Monitoring the Status of Your Instances

You can monitor the status of your instances by viewing status checks and scheduled events for your instances.

A status check gives you the information that results from automated checks performed by Amazon EC2. These automated checks detect whether specific issues are affecting your instances. The status check information, together with the data provided by Amazon CloudWatch, gives you detailed operational visibility into each of your instances.

You can also see status of specific events that are scheduled for your instances. The status of events provides information about upcoming activities that are planned for your instances, such as rebooting or retirement. They also provide the scheduled start and end time of each event.

Contents
• Status Checks for Your Instances (p. 535)
• Scheduled Events for Your Instances (p. 539)

Status Checks for Your Instances

With instance status monitoring, you can quickly determine whether Amazon EC2 has detected any problems that might prevent your instances from running applications. Amazon EC2 performs automated checks on every running EC2 instance to identify hardware and software issues. You can view the results of these status checks to identify specific and detectable problems. The event status data augments the information that Amazon EC2 already provides about the state of each instance (such as pending, running, stopping) and the utilization metrics that Amazon CloudWatch monitors (CPU utilization, network traffic, and disk activity).

Status checks are performed every minute, returning a pass or a fail status. If all checks pass, the overall status of the instance is OK. If one or more checks fail, the overall status is impaired. Status checks are built into Amazon EC2, so they cannot be disabled or deleted.

When a status check fails, the corresponding CloudWatch metric for status checks is incremented. For more information, see Status Check Metrics (p. 553). You can use these metrics to create CloudWatch alarms that are triggered based on the result of the status checks. For example, you can create an alarm to warn you if status checks fail on a specific instance. For more information, see Creating and Editing Status Check Alarms (p. 538).

You can also create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance and automatically recovers the instance if it becomes impaired due to an underlying issue. For more information, see Recover Your Instance (p. 393).
Types of Status Checks

There are two types of status checks: system status checks and instance status checks.

System Status Checks

Monitor the AWS systems on which your instance runs. These checks detect underlying problems with your instance that require AWS involvement to repair. When a system status check fails, you can choose to wait for AWS to fix the issue, or you can resolve it yourself. For instances backed by Amazon EBS, you can stop and start the instance yourself, which in most cases results in the instance being migrated to a new host. For instances backed by instance store, you can terminate and replace the instance.

The following are examples of problems that can cause system status checks to fail:

- Loss of network connectivity
- Loss of system power
- Software issues on the physical host
- Hardware issues on the physical host that impact network reachability

Instance Status Checks

Monitor the software and network configuration of your individual instance. Amazon EC2 checks the health of the instance by sending an address resolution protocol (ARP) request to the network interface (NIC). These checks detect problems that require your involvement to repair. When an instance status check fails, you typically must address the problem yourself (for example, by rebooting the instance or by making instance configuration changes).

The following are examples of problems that can cause instance status checks to fail:

- Failed system status checks
- Incorrect networking or startup configuration
- Exhausted memory
- Corrupted file system
- During instance reboot or while a Windows instance store-backed instance is being bundled, an instance status check reports a failure until the instance becomes available again.

Viewing Status Checks

Amazon EC2 provides you with several ways to view and work with status checks.

Viewing Status Using the Console

You can view status checks using the AWS Management Console.

To view status checks (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. On the **Instances** page, the **Status Checks** column lists the operational status of each instance.

4. To view the status of a specific instance, select the instance, and then choose the **Status Checks** tab. If you have an instance with a failed status check and the instance has been unreachable for over 20 minutes, choose **AWS Support** to submit a request for assistance.

5. To review the CloudWatch metrics for status checks, select the instance, and then choose the **Monitoring** tab. Scroll until you see the graphs for the following metrics:
   - Status Check Failed (Any)
   - Status Check Failed (Instance)
   - Status Check Failed (System)

**Viewing Status Using the Command Line**

You can view status checks for running instances using the describe-instance-status (AWS CLI) command.

To view the status of all instances, use the following command.

```
aws ec2 describe-instance-status
```

To get the status of all instances with an instance status of impaired, use the following command.

```
aws ec2 describe-instance-status --filters Name=instance-status.status,Values=impaired
```

To get the status of a single instance, use the following command.

```
aws ec2 describe-instance-status --instance-ids i-1234567890abcdef0
```

Alternatively, use the following commands:

- **Get-EC2InstanceState** (AWS Tools for Windows PowerShell)
- **DescribeInstanceStatus** (Amazon EC2 Query API)

**Reporting Instance Status**

You can provide feedback if you are having problems with an instance whose status is not shown as impaired, or if you want to send AWS additional details about the problems you are experiencing with an impaired instance.

We use reported feedback to identify issues impacting multiple customers, but do not respond to individual account issues. Providing feedback does not change the status check results that you currently see for the instance.
Reporting Status Feedback Using the Console

To report instance status (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance, choose the Status Checks tab, and choose Submit feedback.
4. Complete the Report Instance Status form, and then choose Submit.

Reporting Status Feedback Using the Command Line

Use the following report-instance-status (AWS CLI) command to send feedback about the status of an impaired instance.

```
aws ec2 report-instance-status --instances i-1234567890abcdef0 --status impaired --reason-codes code
```

Alternatively, use the following commands:

- `Send-EC2InstanceStatus` (AWS Tools for Windows PowerShell)
- `ReportInstanceStatus` (Amazon EC2 Query API)

Creating and Editing Status Check Alarms

You can use the status check metrics (p. 553) to create CloudWatch alarms to notify you when an instance has a failed status check.

Creating a Status Check Alarm Using the Console

Use the following procedure to configure an alarm that sends you a notification by email, or stops, terminates, or recovers an instance when it fails a status check.

To create a status check alarm (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance, choose the Status Checks tab, and choose Create Status Check Alarm.
4. Select Send a notification to. Choose an existing SNS topic, or choose create topic to create a new one. If creating a new topic, in With these recipients, enter your email address and the addresses of any additional recipients, separated by commas.
5. (Optional) Select Take the action, and then select the action that you’d like to take.
6. In Whenever, select the status check that you want to be notified about.
7. If you selected Recover this instance in the previous step, select Status Check Failed (System).
8. In For at least, set the number of periods you want to evaluate and in consecutive periods, select the evaluation period duration before triggering the alarm and sending an email.
9. (Optional) In Name of alarm, replace the default name with another name for the alarm.
10. Choose Create Alarm.

**Important**

If you added an email address to the list of recipients or created a new topic, Amazon SNS sends a subscription confirmation email message to each new address. Each recipient must confirm the subscription by choosing the link contained in that message. Alert notifications are sent only to confirmed addresses.
If you need to make changes to an instance status alarm, you can edit it.

To edit a status check alarm using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Select the instance and choose **Actions**, **CloudWatch Monitoring, Add/Edit Alarms**.
4. In the **Alarm Details** dialog box, choose the name of the alarm.
5. In the **Edit Alarm** dialog box, make the desired changes, and then choose **Save**.

Creating a Status Check Alarm Using the AWS CLI

In the following example, the alarm publishes a notification to an SNS topic, arn:aws:sns:us-west-2:111122223333:my-sns-topic, when the instance fails either the instance check or system status check for at least two consecutive periods. The CloudWatch metric used is StatusCheckFailed.

To create a status check alarm using the AWS CLI

1. Select an existing SNS topic or create a new one. For more information, see Using the AWS CLI with Amazon SNS in the **AWS Command Line Interface User Guide**.
2. Use the following list-metrics command to view the available Amazon CloudWatch metrics for Amazon EC2.

   ```
   aws cloudwatch list-metrics --namespace AWS/EC2
   ```

3. Use the following put-metric-alarm command to create the alarm.

   ```
   aws cloudwatch put-metric-alarm --alarm-name StatusCheckFailed-Alarm-for-i-1234567890abcdef0 --metric-name StatusCheckFailed --namespace AWS/EC2 --statistic Maximum --dimensions Name=InstanceId,Value=i-1234567890abcdef0 --unit Count --period 300 --evaluation-periods 2 --threshold 1 --comparison-operator GreaterThanOrEqualToThreshold --alarm-actions arn:aws:sns:us-west-2:111122223333:my-sns-topic
   ```

The period is the time frame, in seconds, in which Amazon CloudWatch metrics are collected. This example uses 300, which is 60 seconds multiplied by 5 minutes. The evaluation period is the number of consecutive periods for which the value of the metric must be compared to the threshold. This example uses 2. The alarm actions are the actions to perform when this alarm is triggered. This example configures the alarm to send an email using Amazon SNS.

Scheduled Events for Your Instances

AWS can schedule events for your instances, such as a reboot, stop/start, or retirement. These events do not occur frequently. If one of your instances will be affected by a scheduled event, AWS sends an email to the email address that's associated with your AWS account prior to the scheduled event. The email provides details about the event, including the start and end date. Depending on the event, you might be able to take action to control the timing of the event.

To update the contact information for your account so that you can be sure to be notified about scheduled events, go to the **Account Settings** page.
• Viewing Scheduled Events (p. 540)
• Working with Instances Scheduled to Stop or Retire (p. 542)
• Working with Instances Scheduled for Reboot (p. 543)
• Working with Instances Scheduled for Maintenance (p. 545)

Types of Scheduled Events

Amazon EC2 supports the following types of scheduled events for your instances:

• **Instance stop**: At the scheduled time, the instance is stopped. When you start it again, it's migrated to a new host. Applies only to instances backed by Amazon EBS.

• **Instance retirement**: At the scheduled time, the instance is stopped if it is backed by Amazon EBS, or terminated if it is backed by instance store.

• **Instance reboot**: At the scheduled time, the instance is rebooted.

• **System reboot**: At the scheduled time, the host for the instance is rebooted.

• **System maintenance**: At the scheduled time, the instance might be temporarily affected by network maintenance or power maintenance.

Viewing Scheduled Events

In addition to receiving notification of scheduled events in email, you can check for scheduled events using one of the following methods.

**To view scheduled events for your instances using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Events**. Any resources with an associated event are displayed. You can filter by resource type, or by specific event types. You can select the resource to view details.

3. Alternatively, in the navigation pane, choose **EC2 Dashboard**. Any resources with an associated event are displayed under **Scheduled Events**.

4. Some events are also shown for affected resources. For example, in the navigation pane, choose **Instances** and select an instance. If the instance has an associated instance stop or instance retirement event, it is displayed in the lower pane.
To view scheduled events for your instances using the AWS CLI

- Use the following `describe-instance-status` command.

```bash
aws ec2 describe-instance-status --instance-id i-1234567890abcdef0 --query "InstanceStatuses[].Events"
```

The following example output shows a reboot event.

```
[
  "Events": [
    {
      "InstanceEventId": "instance-event-0d59937288b749b32",
      "Code": "system-reboot",
      "Description": "The instance is scheduled for a reboot",
      "NotAfter": "2019-03-15T22:00:00.000Z",
      "NotBefore": "2019-03-14T20:00:00.000Z",
      "NotBeforeDeadline": "2019-04-05T11:00:00.000Z"
    }
  ]
]
```

The following example output shows an instance retirement event.

```
[
  "Events": [
    {
      "InstanceEventId": "instance-event-0e439355b779n26",
      "Code": "instance-stop",
      "Description": "The instance is running on degraded hardware",
      "NotBefore": "2015-05-23T00:00:00.000Z"
    }
  ]
]
```

To view scheduled events for your instances using the AWS Tools for Windows PowerShell

- Use the following `Get-EC2InstanceStatus` command.

```powershell
PS C:\> (Get-EC2InstanceStatus -InstanceId i-1234567890abcdef0).Events
```

The following example output shows an instance retirement event.

```
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>NotBefore</th>
</tr>
</thead>
<tbody>
<tr>
<td>instance-stop</td>
<td>The instance is running on degraded hardware</td>
<td>5/23/2015 12:00:00 AM</td>
</tr>
</tbody>
</table>
```
To view scheduled events for your instances using instance metadata

- You can retrieve information about active maintenance events for your instances from the instance metadata (p. 478), as follows.

```
```

The following is example output with information about a scheduled system reboot event, in JSON format.

```json
[
  {
    "NotBefore": "21 Jan 2019 09:00:43 GMT",
    "Code": "system-reboot",
    "Description": "scheduled reboot",
    "EventId": "instance-event-0d59937288b749b32",
    "NotAfter": "21 Jan 2019 09:17:23 GMT",
    "State": "active"
  }
]
```

To view event history about completed or canceled events for your instances using instance metadata

- You can retrieve information about completed or canceled events for your instances from the instance metadata (p. 478), as follows.

```
```

The following is example output with information about a system reboot event that was canceled, and a system reboot event that was completed, in JSON format.

```json
[
  {
    "NotBefore": "21 Jan 2019 09:00:43 GMT",
    "Code": "system-reboot",
    "Description": "[Canceled] scheduled reboot",
    "EventId": "instance-event-0d59937288b749b32",
    "NotAfter": "21 Jan 2019 09:17:23 GMT",
    "State": "canceled"
  },
  {
    "NotBefore": "29 Jan 2019 09:00:43 GMT",
    "Code": "system-reboot",
    "Description": "[Completed] scheduled reboot",
    "EventId": "instance-event-0d59937288b749b32",
    "NotAfter": "29 Jan 2019 09:17:23 GMT",
    "State": "completed"
  }
]
```

**Working with Instances Scheduled to Stop or Retire**

When AWS detects irreparable failure of the underlying host for your instance, it schedules the instance to stop or terminate, depending on the type of root device for the instance. If the root device is an EBS
volume, the instance is scheduled to stop. If the root device is an instance store volume, the instance is scheduled to terminate. For more information, see Instance Retirement (p. 386).

**Important**
Any data stored on instance store volumes is lost when an instance is stopped or terminated. This includes instance store volumes that are attached to an instance that has an EBS volume as the root device. Be sure to save data from your instance store volumes that you might need later before the instance is stopped or terminated.

**Actions for Instances Backed by Amazon EBS**
You can wait for the instance to stop as scheduled. Alternatively, you can stop and start the instance yourself, which migrates it to a new host. For more information about stopping your instance, in addition to information about the changes to your instance configuration when it's stopped, see Stop and Start Your Instance (p. 383).

You can automate an immediate stop and start in response to a scheduled instance stop event. For more information, see Automating Actions for EC2 Instances in the AWS Health User Guide.

**Actions for Instances Backed by Instance Store**
We recommend that you launch a replacement instance from your most recent AMI and migrate all necessary data to the replacement instance before the instance is scheduled to terminate. Then, you can terminate the original instance, or wait for it to terminate as scheduled.

**Working with Instances Scheduled for Reboot**
When AWS must perform tasks such as installing updates or maintaining the underlying host, it can schedule the instance or the underlying host for a reboot. You can reschedule most reboot events (p. 544) so that your instance is rebooted at a specific date and time that suits you.

If you stop your linked EC2-Classic instance (p. 763), it is automatically unlinked from the VPC and the VPC security groups are no longer associated with the instance. You can link your instance to the VPC again after you've restarted it.

**Viewing the Reboot Event Type**
You can view whether a reboot event is an instance reboot or a system reboot using the AWS Management Console, AWS CLI, or Amazon EC2 API.

**To view the type of scheduled reboot event (console)**
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Events.
3. Choose Instance resources from the filter list.
4. For each instance, view the value in the Event Type column. The value is either system-reboot or instance-reboot.

**To view the type of scheduled reboot event (AWS CLI)**
- Use the following describe-instance-status command.

```
aws ec2 describe-instance-status --instance-id i-1234567890abcdef0
```

For scheduled reboot events, the value for Code is either system-reboot or instance-reboot. The following example output shows a system-reboot event.

```
[
]
```
Actions for Instance Reboot

You can wait for the instance reboot to occur within its scheduled maintenance window, reschedule (p. 544) the instance reboot to a date and time that suits you, or reboot (p. 385) the instance yourself at a time that is convenient for you.

After your instance is rebooted, the scheduled event is cleared and the event’s description is updated. The pending maintenance to the underlying host is completed, and you can begin using your instance again after it has fully booted.

Actions for System Reboot

It is not possible for you to reboot the system yourself. You can wait for the system reboot to occur during its scheduled maintenance window, or you can reschedule (p. 544) the system reboot to a date and time that suits you. A system reboot typically completes in a matter of minutes. After the system reboot has occurred, the instance retains its IP address and DNS name, and any data on local instance store volumes is preserved. After the system reboot is complete, the scheduled event for the instance is cleared, and you can verify that the software on your instance is operating as expected.

Alternatively, if it is necessary to maintain the instance at a different time and you can't reschedule the system reboot, then you can stop and start an Amazon EBS-backed instance, which migrates it to a new host. However, the data on the local instance store volumes is not preserved. You can also automate an immediate instance stop and start in response to a scheduled system reboot event. For more information, see Automating Actions for EC2 Instances in the AWS Health User Guide. For an instance store-backed instance, if you can't reschedule the system reboot, then you can launch a replacement instance from your most recent AMI, migrate all necessary data to the replacement instance before the scheduled maintenance window, and then terminate the original instance.

Rescheduling a Reboot Event

You can reschedule most reboot events so that your instance is rebooted at a specific date and time that suits you.

To reschedule a reboot event (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Events.
3. Choose Instance resources from the filter list.
4. Select one or more instances, and then choose Actions, Schedule Event.

Only events that have an event deadline date, indicated by a value for Event Deadline, can be rescheduled.

5. For Event start time, enter a new date and time for the reboot. The new date and time must fall before the Event Deadline.

It might take 1-2 minutes for the updated event start time to be reflected in the console.
To reschedule a reboot event (AWS CLI)

1. Only events that have an event deadline date, indicated by a value for `NotBeforeDeadline`, can be rescheduled. Use the following `describe-instance-status` command to view the `NotBeforeDeadline` parameter value.

   ```bash
   aws ec2 describe-instance-status --instance-id i-1234567890abcdef0
   ```

   The following example output shows a `system-reboot` event that can be rescheduled because `NotBeforeDeadline` contains a value.

   ```json
   [  
   "Events": [ 
   
   "InstanceEventId": "instance-event-0d9937288b749b32", 
   "Code": "system-reboot", 
   "Description": "The instance is scheduled for a reboot", 
   "NotAfter": "2019-03-14T22:00:00.000Z", 
   "NotBefore": "2019-03-14T20:00:00.000Z", 
   "NotBeforeDeadline": "2019-04-05T11:00:00.000Z"
   } 
   ]
   ```

2. To reschedule the event, use the `modify-instance-event-start-time` command. Specify the new event start time using the `not-before` parameter. The new event start time must fall before the `NotBeforeDeadline`.

   ```bash
   aws ec2 modify-instance-event-start-time --instance-id i-1234567890abcdef0 
   --instance-event-id instance-event-0d9937288b749b32 
   --not-before 2019-03-25T10:00:00.000
   ```

   It might take 1-2 minutes before the `describe-instance-status` command returns the updated `NotBefore` parameter value.

Limitations for Reboot Events

- Only reboot events with an event deadline date can be rescheduled. The event can be rescheduled up to the event deadline date. The `Event Deadline` column in the console and the `NotBeforeDeadline` field in the AWS CLI indicate if the event has a deadline date.
- Only reboot events that have not yet started can be rescheduled. The `Start Time` column in the console and the `NotBefore` field in the AWS CLI indicate the event start time. Reboot events that are scheduled to start in the next 5 minutes cannot be rescheduled.
- The new event start time must be at least 60 minutes from the current time.
- If you reschedule multiple events using the console, the event deadline date is determined by the event with earliest event deadline date.

Working with Instances Scheduled for Maintenance

When AWS must maintain the underlying host for an instance, it schedules the instance for maintenance. There are two types of maintenance events: network maintenance and power maintenance.

During network maintenance, scheduled instances lose network connectivity for a brief period of time. Normal network connectivity to your instance will be restored after maintenance is complete.

During power maintenance, scheduled instances are taken offline for a brief period, and then rebooted. When a reboot is performed, all of your instance's configuration settings are retained.
After your instance has rebooted (this normally takes a few minutes), verify that your application is working as expected. At this point, your instance should no longer have a scheduled event associated with it, or the description of the scheduled event begins with [Completed]. It sometimes takes up to 1 hour for the instance status description to refresh. Completed maintenance events are displayed on the Amazon EC2 console dashboard for up to a week.

### Actions for Instances Backed by Amazon EBS

You can wait for the maintenance to occur as scheduled. Alternatively, you can stop and start the instance, which migrates it to a new host. For more information about stopping your instance, in addition to information about the changes to your instance configuration when it's stopped, see Stop and Start Your Instance (p. 383).

You can automate an immediate stop and start in response to a scheduled maintenance event. For more information, see Automating Actions for EC2 Instances in the AWS Health User Guide.

### Actions for Instances Backed by Instance Store

You can wait for the maintenance to occur as scheduled. Alternatively, if you want to maintain normal operation during a scheduled maintenance window, you can launch a replacement instance from your most recent AMI, migrate all necessary data to the replacement instance before the scheduled maintenance window, and then terminate the original instance.

---

## Monitoring Your Instances Using CloudWatch

You can monitor your instances using Amazon CloudWatch, which collects and processes raw data from Amazon EC2 into readable, near real-time metrics. These statistics are recorded for a period of 15 months, so that you can access historical information and gain a better perspective on how your web application or service is performing.

By default, Amazon EC2 sends metric data to CloudWatch in 5-minute periods. To send metric data for your instance to CloudWatch in 1-minute periods, you can enable detailed monitoring on the instance. For more information, see Enable or Disable Detailed Monitoring for Your Instances (p. 546).

The Amazon EC2 console displays a series of graphs based on the raw data from Amazon CloudWatch. Depending on your needs, you might prefer to get data for your instances from Amazon CloudWatch instead of the graphs in the console.

For more information about Amazon CloudWatch, see the Amazon CloudWatch User Guide.

### Contents

- Enable or Disable Detailed Monitoring for Your Instances (p. 546)
- List the Available CloudWatch Metrics for Your Instances (p. 548)
- Get Statistics for Metrics for Your Instances (p. 557)
- Graph Metrics for Your Instances (p. 564)
- Create a CloudWatch Alarm for an Instance (p. 564)
- Create Alarms That Stop, Terminate, Reboot, or Recover an Instance (p. 565)

---

## Enable or Disable Detailed Monitoring for Your Instances

By default, your instance is enabled for basic monitoring. You can optionally enable detailed monitoring. After you enable detailed monitoring, the Amazon EC2 console displays monitoring graphs with a 1-minute period for the instance. The following table describes basic and detailed monitoring for instances.
### Monitoring Type

<table>
<thead>
<tr>
<th>Monitoring Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Data is available automatically in 5-minute periods at no charge.</td>
</tr>
<tr>
<td>Detailed</td>
<td>Data is available in 1-minute periods for an additional cost. To get this level of data, you must specifically enable it for the instance. For the instances where you've enabled detailed monitoring, you can also get aggregated data across groups of similar instances. For information about pricing, see the Amazon CloudWatch product page.</td>
</tr>
</tbody>
</table>

### Enabling Detailed Monitoring

You can enable detailed monitoring on an instance as you launch it or after the instance is running or stopped. Enabling detailed monitoring on an instance does not affect the monitoring of the EBS volumes attached to the instance. For more information, see Amazon CloudWatch Metrics for Amazon EBS (p. 903).

**To enable detailed monitoring for an existing instance (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Select the instance and choose **Actions, CloudWatch Monitoring, Enable Detailed Monitoring**.
4. In the **Enable Detailed Monitoring** dialog box, choose **Yes, Enable**.
5. Choose **Close**.

**To enable detailed monitoring when launching an instance (console)**

When launching an instance using the AWS Management Console, select the **Monitoring** check box on the **Configure Instance Details** page.

**To enable detailed monitoring for an existing instance (AWS CLI)**

Use the following `monitor-instances` command to enable detailed monitoring for the specified instances.

```
aws ec2 monitor-instances --instance-ids i-1234567890abcdef0
```

**To enable detailed monitoring when launching an instance (AWS CLI)**

Use the `run-instances` command with the `--monitoring` flag to enable detailed monitoring.

```
aws ec2 run-instances --image-id ami-09092360 --monitoring Enabled=true...
```

### Disabling Detailed Monitoring

You can disable detailed monitoring on an instance as you launch it or after the instance is running or stopped.

**To disable detailed monitoring (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Select the instance and choose **Actions, CloudWatch Monitoring, Disable Detailed Monitoring**.
4. In the **Disable Detailed Monitoring** dialog box, choose **Yes, Disable**.
5. Choose **Close**.

**To disable detailed monitoring (AWS CLI)**

Use the following `unmonitor-instances` command to disable detailed monitoring for the specified instances.

```
aws ec2 unmonitor-instances --instance-ids i-1234567890abcdef0
```

**List the Available CloudWatch Metrics for Your Instances**

Amazon EC2 sends metrics to Amazon CloudWatch. You can use the AWS Management Console, the AWS CLI, or an API to list the metrics that Amazon EC2 sends to CloudWatch. By default, each data point covers the 5 minutes that follow the start time of activity for the instance. If you've enabled detailed monitoring, each data point covers the next minute of activity from the start time.

For information about getting the statistics for these metrics, see Get Statistics for Metrics for Your Instances (p. 557).

**Contents**

- **Instance Metrics** (p. 548)
- **CPU Credit Metrics** (p. 550)
- **Amazon EBS Metrics for Nitro-based Instances** (p. 552)
- **Status Check Metrics** (p. 553)
- **Amazon EC2 Metric Dimensions** (p. 554)
- **Listing Metrics Using the Console** (p. 554)
- **Listing Metrics Using the AWS CLI** (p. 556)

**Instance Metrics**

The AWS/EC2 namespace includes the following instance metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUUtilization</td>
<td>The percentage of allocated EC2 compute units that are currently in use on the instance. This metric identifies the processing power required to run an application upon a selected instance. Depending on the instance type, tools in your operating system can show a lower percentage than CloudWatch when the instance is not allocated a full processor core. Units: Percent</td>
</tr>
<tr>
<td>DiskReadOps</td>
<td>Completed read operations from all instance store volumes available to the instance in a specified period of time.</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>To calculate the average I/O operations per second (IOPS) for the period, divide the total operations in the period by the number of seconds in that period.</td>
</tr>
<tr>
<td></td>
<td>If there are no instance store volumes, either the value is 0 or the metric is not reported.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>DiskWriteOps</td>
<td>Completed write operations to all instance store volumes available to the instance in a specified period of time.</td>
</tr>
<tr>
<td></td>
<td>To calculate the average I/O operations per second (IOPS) for the period, divide the total operations in the period by the number of seconds in that period.</td>
</tr>
<tr>
<td></td>
<td>If there are no instance store volumes, either the value is 0 or the metric is not reported.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
<tr>
<td>DiskReadBytes</td>
<td>Bytes read from all instance store volumes available to the instance.</td>
</tr>
<tr>
<td></td>
<td>This metric is used to determine the volume of the data the application reads from the hard disk of the instance. This can be used to determine the speed of the application.</td>
</tr>
<tr>
<td></td>
<td>The number reported is the number of bytes received during the period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to find Bytes/second. If you have detailed (one-minute) monitoring, divide it by 60.</td>
</tr>
<tr>
<td></td>
<td>If there are no instance store volumes, either the value is 0 or the metric is not reported.</td>
</tr>
<tr>
<td></td>
<td>Units: Bytes</td>
</tr>
<tr>
<td>DiskWriteBytes</td>
<td>Bytes written to all instance store volumes available to the instance.</td>
</tr>
<tr>
<td></td>
<td>This metric is used to determine the volume of the data the application writes onto the hard disk of the instance. This can be used to determine the speed of the application.</td>
</tr>
<tr>
<td></td>
<td>The number reported is the number of bytes received during the period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to find Bytes/second. If you have detailed (one-minute) monitoring, divide it by 60.</td>
</tr>
<tr>
<td></td>
<td>If there are no instance store volumes, either the value is 0 or the metric is not reported.</td>
</tr>
<tr>
<td></td>
<td>Units: Bytes</td>
</tr>
</tbody>
</table>
Amazon Elastic Compute Cloud  
User Guide for Windows Instances  
List Available Metrics

### Metric | Description
---|---
**NetworkIn** | The number of bytes received on all network interfaces by the instance. This metric identifies the volume of incoming network traffic to a single instance.

The number reported is the number of bytes received during the period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to find Bytes/second. If you have detailed (one-minute) monitoring, divide it by 60.

Units: Bytes

**NetworkOut** | The number of bytes sent out on all network interfaces by the instance. This metric identifies the volume of outgoing network traffic from a single instance.

The number reported is the number of bytes sent during the period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to find Bytes/second. If you have detailed (one-minute) monitoring, divide it by 60.

Units: Bytes

**NetworkPacketsIn** | The number of packets received on all network interfaces by the instance. This metric identifies the volume of incoming traffic in terms of the number of packets on a single instance. This metric is available for basic monitoring only.

Units: Count

Statistics: Minimum, Maximum, Average

**NetworkPacketsOut** | The number of packets sent out on all network interfaces by the instance. This metric identifies the volume of outgoing traffic in terms of the number of packets on a single instance. This metric is available for basic monitoring only.

Units: Count

Statistics: Minimum, Maximum, Average

---

**CPU Credit Metrics**

The AWS/EC2 namespace includes the following CPU credit metrics for your burstable performance instances (p. 132).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CPUCreditUsage</strong></td>
<td>The number of CPU credits spent by the instance for CPU utilization. One CPU credit equals one vCPU running at 100% utilization for one minute or an equivalent combination of vCPUs, utilization, and time (for example, one vCPU running at 50% utilization for two minutes or two vCPUs running at 25% utilization for two minutes).</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>CPU credit metrics are available at a five-minute frequency only. If you specify a period greater than five minutes, use the Sum statistic instead of the Average statistic.</td>
</tr>
<tr>
<td></td>
<td>Units: Credits (vCPU-minutes)</td>
</tr>
<tr>
<td></td>
<td>The number of earned CPU credits that an instance has accrued since it was launched or started. For T2 Standard, the CPUCreditBalance also includes the number of launch credits that have been accrued.</td>
</tr>
<tr>
<td>CPUCreditBalance</td>
<td>Credits are accrued in the credit balance after they are earned, and removed from the credit balance when they are spent. The credit balance has a maximum limit, determined by the instance size. After the limit is reached, any new credits that are earned are discarded. For T2 Standard, launch credits do not count towards the limit.</td>
</tr>
<tr>
<td></td>
<td>The credits in the CPUCreditBalance are available for the instance to spend to burst beyond its baseline CPU utilization.</td>
</tr>
<tr>
<td></td>
<td>When an instance is running, credits in the CPUCreditBalance do not expire. When a T3 or T3a instance stops, the CPUCreditBalance value persists for seven days. Thereafter, all accrued credits are lost. When a T2 instance stops, the CPUCreditBalance value does not persist, and all accrued credits are lost.</td>
</tr>
<tr>
<td></td>
<td>CPU credit metrics are available at a five-minute frequency only.</td>
</tr>
<tr>
<td>CPUSurplusCreditBalance</td>
<td>Units: Credits (vCPU-minutes)</td>
</tr>
<tr>
<td></td>
<td>The number of surplus credits that have been spent by an unlimited instance when its CPUCreditBalance value is zero.</td>
</tr>
<tr>
<td></td>
<td>The CPUSurplusCreditBalance value is paid down by earned CPU credits. If the number of surplus credits exceeds the maximum number of credits that the instance can earn in a 24-hour period, the spent surplus credits above the maximum incur an additional charge.</td>
</tr>
<tr>
<td>CPUSurplusCreditsCharged</td>
<td>The number of spent surplus credits that are not paid down by earned CPU credits, and which thus incur an additional charge.</td>
</tr>
<tr>
<td></td>
<td>Spent surplus credits are charged when any of the following occurs:</td>
</tr>
<tr>
<td></td>
<td>• The spent surplus credits exceed the maximum number of credits that the instance can earn in a 24-hour period. Spent surplus credits above the maximum are charged at the end of the hour.</td>
</tr>
<tr>
<td></td>
<td>• The instance is stopped or terminated.</td>
</tr>
<tr>
<td></td>
<td>• The instance is switched from unlimited to standard.</td>
</tr>
<tr>
<td>CPUSurplusCreditsCharged</td>
<td>Units: Credits (vCPU-minutes)</td>
</tr>
</tbody>
</table>
Amazon EBS Metrics for Nitro-based Instances

The AWS/EC2 namespace includes the following Amazon EBS metrics for the Nitro-based instances that are not bare metal instances. For the list of Nitro-based instance types, see Nitro-based Instances (p. 122).

Metric values for Nitro-based instances will always be integers (whole numbers), whereas values for Xen-based instances support decimals. Therefore, low instance CPU utilization on Nitro-based instances may appear to be rounded down to 0.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSReadOps</td>
<td>Completed read operations from all Amazon EBS volumes attached to the instance in a specified period of time.</td>
</tr>
<tr>
<td></td>
<td>To calculate the average read I/O operations per second (Read IOPS) for the period, divide the total operations in the period by the number of seconds in that period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to calculate the Read IOPS. If you have detailed (one-minute) monitoring, divide it by 60.</td>
</tr>
<tr>
<td></td>
<td>Unit: Count</td>
</tr>
<tr>
<td>EBSWriteOps</td>
<td>Completed write operations to all EBS volumes attached to the instance in a specified period of time.</td>
</tr>
<tr>
<td></td>
<td>To calculate the average write I/O operations per second (Write IOPS) for the period, divide the total operations in the period by the number of seconds in that period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to calculate the Write IOPS. If you have detailed (one-minute) monitoring, divide it by 60.</td>
</tr>
<tr>
<td></td>
<td>Unit: Count</td>
</tr>
<tr>
<td>EBSReadBytes</td>
<td>Bytes read from all EBS volumes attached to the instance in a specified period of time.</td>
</tr>
<tr>
<td></td>
<td>The number reported is the number of bytes read during the period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to find Read Bytes/second. If you have detailed (one-minute) monitoring, divide it by 60.</td>
</tr>
<tr>
<td></td>
<td>Unit: Bytes</td>
</tr>
<tr>
<td>EBSWriteBytes</td>
<td>Bytes written to all EBS volumes attached to the instance in a specified period of time.</td>
</tr>
<tr>
<td></td>
<td>The number reported is the number of bytes written during the period. If you are using basic (five-minute) monitoring, you can divide this number by 300 to find Write Bytes/second. If you have detailed (one-minute) monitoring, divide it by 60.</td>
</tr>
</tbody>
</table>
### List Available Metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBSIODBalance%</td>
<td>Available only for the smaller instance sizes. Provides information about the percentage of I/O credits remaining in the burst bucket. This metric is available for basic monitoring only. The <strong>Sum</strong> statistic is not applicable to this metric. Unit: Percent</td>
</tr>
<tr>
<td>EBSByteBalance%</td>
<td>Available only for the smaller instance sizes. Provides information about the percentage of throughput credits remaining in the burst bucket. This metric is available for basic monitoring only. The <strong>Sum</strong> statistic is not applicable to this metric. Unit: Percent</td>
</tr>
</tbody>
</table>

For information about the metrics provided for your EBS volumes, see Amazon EBS Metrics (p. 903). For information about the metrics provided for your Spot fleets, see CloudWatch Metrics for Spot Fleet (p. 272).

### Status Check Metrics

The **AWS/EC2** namespace includes the following status check metrics. By default, status check metrics are available at a 1-minute frequency at no charge. For a newly-launched instance, status check metric data is only available after the instance has completed the initialization state (within a few minutes of the instance entering the running state). For more information about EC2 status checks, see Status Checks For Your Instances.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>StatusCheckFailed</td>
<td>Reports whether the instance has passed both the instance status check and the system status check in the last minute. This metric can be either 0 (passed) or 1 (failed). By default, this metric is available at a 1-minute frequency at no charge. Units: Count</td>
</tr>
<tr>
<td>StatusCheckFailed_Instance</td>
<td>Reports whether the instance has passed the instance status check in the last minute. This metric can be either 0 (passed) or 1 (failed). By default, this metric is available at a 1-minute frequency at no charge. Units: Count</td>
</tr>
<tr>
<td>StatusCheckFailed_System</td>
<td>Reports whether the instance has passed the system status check in the last minute.</td>
</tr>
</tbody>
</table>
Amazon Elastic Compute Cloud
User Guide for Windows Instances
List Available Metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This metric can be either 0 (passed) or 1 (failed).</td>
</tr>
<tr>
<td></td>
<td>By default, this metric is available at a 1-minute frequency at no charge.</td>
</tr>
<tr>
<td></td>
<td>Units: Count</td>
</tr>
</tbody>
</table>

Amazon EC2 Metric Dimensions

You can use the following dimensions to refine the metrics returned for your instances.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AutoScalingGroupName</td>
<td>This dimension filters the data you request for all instances in a specified capacity group. An Auto Scaling group is a collection of instances you define if you’re using Auto Scaling. This dimension is available only for Amazon EC2 metrics when the instances are in such an Auto Scaling group. Available for instances with Detailed or Basic Monitoring enabled.</td>
</tr>
<tr>
<td>ImageId</td>
<td>This dimension filters the data you request for all instances running this Amazon EC2 Amazon Machine Image (AMI). Available for instances with Detailed Monitoring enabled.</td>
</tr>
<tr>
<td>InstanceId</td>
<td>This dimension filters the data you request for the identified instance only. This helps you pinpoint an exact instance from which to monitor data.</td>
</tr>
<tr>
<td>InstanceType</td>
<td>This dimension filters the data you request for all instances running with this specified instance type. This helps you categorize your data by the type of instance running. For example, you might compare data from an m1.small instance and an m1.large instance to determine which has the better business value for your application. Available for instances with Detailed Monitoring enabled.</td>
</tr>
</tbody>
</table>

Listing Metrics Using the Console

Metrics are grouped first by namespace, and then by the various dimension combinations within each namespace. For example, you can view all metrics provided by Amazon EC2, or metrics grouped by instance ID, instance type, image (AMI) ID, or Auto Scaling group.

**To view available metrics by category (console)**

2. In the navigation pane, choose Metrics.
3. Choose the EC2 metric namespace.
4. Select a metric dimension (for example, **Per-Instance Metrics**).

5. To sort the metrics, use the column heading. To graph a metric, select the check box next to the metric. To filter by resource, choose the resource ID and then choose **Add to search**. To filter by metric, choose the metric name and then choose **Add to search**.
Listing Metrics Using the AWS CLI

Use the `list-metrics` command to list the CloudWatch metrics for your instances.

To list all the available metrics for Amazon EC2 (AWS CLI)

The following example specifies the `AWS/EC2` namespace to view all the metrics for Amazon EC2.

```
aws cloudwatch list-metrics --namespace AWS/EC2
```

The following is example output:

```
{
  "Metrics": [
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "NetworkOut"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "CPUUtilization"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "DiskReadBytes"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "DiskWriteBytes"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "DiskReadOps"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "DiskWriteOps"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "NetworkIn"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "NetworkOut"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "DiskReadOps"
    },
    {
      "Namespace": "AWS/EC2",
      "Dimensions": [
        {
          "Name": "InstanceId",
          "Value": "i-1234567890abcdef0"
        }
      ],
      "MetricName": "DiskWriteOps"
    }
  ]
}
```
To list all the available metrics for an instance (AWS CLI)

The following example specifies the AWS/EC2 namespace and the InstanceId dimension to view the results for the specified instance only.

```
aws cloudwatch list-metrics --namespace AWS/EC2 --dimensions
  Name=InstanceId,Value=i-1234567890abcdef0
```

To list a metric across all instances (AWS CLI)

The following example specifies the AWS/EC2 namespace and a metric name to view the results for the specified metric only.

```
aws cloudwatch list-metrics --namespace AWS/EC2 --metric-name CPUUtilization
```

Get Statistics for Metrics for Your Instances

You can get statistics for the CloudWatch metrics for your instances.

Contents
- Statistics Overview (p. 557)
- Get Statistics for a Specific Instance (p. 558)
- Aggregate Statistics Across Instances (p. 560)
- Aggregate Statistics by Auto Scaling Group (p. 562)
- Aggregate Statistics by AMI (p. 563)

Statistics Overview

*Statistics* are metric data aggregations over specified periods of time. CloudWatch provides statistics based on the metric data points provided by your custom data or provided by other services in AWS to CloudWatch. Aggregations are made using the namespace, metric name, dimensions, and the data point unit of measure, within the time period you specify. The following table describes the available statistics.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>The lowest value observed during the specified period. You can use this value to determine low volumes of activity for your application.</td>
</tr>
<tr>
<td>Maximum</td>
<td>The highest value observed during the specified period. You can use this value to determine high volumes of activity for your application.</td>
</tr>
<tr>
<td>Sum</td>
<td>All values submitted for the matching metric added together. This statistic can be useful for determining the total volume of a metric.</td>
</tr>
<tr>
<td>Average</td>
<td>The value of Sum / SampleCount during the specified period. By comparing this statistic with the Minimum and Maximum, you can determine the full scope of a metric.</td>
</tr>
<tr>
<td>Statistic</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>and how close the average use is to the Minimum and Maximum. This comparison helps you to know when to increase or decrease your resources as needed.</td>
</tr>
<tr>
<td>SampleCount</td>
<td>The count (number) of data points used for the statistical calculation.</td>
</tr>
<tr>
<td>pNN.NN</td>
<td>The value of the specified percentile. You can specify any percentile, using up to two decimal places (for example, p95.45).</td>
</tr>
</tbody>
</table>

Get Statistics for a Specific Instance

The following examples show you how to use the AWS Management Console or the AWS CLI to determine the maximum CPU utilization of a specific EC2 instance.

Requirements

- You must have the ID of the instance. You can get the instance ID using the AWS Management Console or the describe-instances command.
- By default, basic monitoring is enabled, but you can enable detailed monitoring. For more information, see Enable or Disable Detailed Monitoring for Your Instances (p. 546).

To display the CPU utilization for a specific instance (console)

2. In the navigation pane, choose Metrics.
3. Choose the EC2 metric namespace.
4. Choose the Per-Instance Metrics dimension.
5. In the search field, enter **CPUUtilization** and press Enter. Choose the row for the specific instance, which displays a graph for the **CPUUtilization** metric for the instance. To name the graph, choose the pencil icon. To change the time range, select one of the predefined values or choose **custom**.

6. To change the statistic or the period for the metric, choose the **Graphed metrics** tab. Choose the column heading or an individual value, and then choose a different value.
To get the CPU utilization for a specific instance (AWS CLI)

Use the following `get-metric-statistics` command to get the **CPUUtilization** metric for the specified instance, using the specified period and time interval:

```
aws cloudwatch get-metric-statistics --namespace AWS/EC2 --metric-name CPUUtilization --period 3600 --statistics Maximum --dimensions Name=InstanceId,Value=i-1234567890abcdef0 --start-time 2016-10-18T23:18:00Z --end-time 2016-10-19T23:18:00Z
```

The following is example output. Each value represents the maximum CPU utilization percentage for a single EC2 instance.

```
{
   "Datapoints": [
     {
       "Timestamp": "2016-10-19T00:18:00Z",
       "Maximum": 0.33000000000000002,
       "Unit": "Percent"
     },
     {
       "Timestamp": "2016-10-19T03:18:00Z",
       "Maximum": 99.670000000000002,
       "Unit": "Percent"
     },
     {
       "Timestamp": "2016-10-19T07:18:00Z",
       "Maximum": 0.34000000000000002,
       "Unit": "Percent"
     },
     {
       "Timestamp": "2016-10-19T12:18:00Z",
       "Maximum": 0.34000000000000002,
       "Unit": "Percent"
     },
     ...
  ],
  "Label": "CPUUtilization"
}
```

Aggregate Statistics Across Instances

Aggregate statistics are available for the instances that have detailed monitoring enabled. Instances that use basic monitoring are not included in the aggregates. In addition, Amazon CloudWatch does not aggregate data across regions. Therefore, metrics are completely separate between regions. Before you
can get statistics aggregated across instances, you must enable detailed monitoring (at an additional charge), which provides data in 1-minute periods.

This example shows you how to use detailed monitoring to get the average CPU usage for your EC2 instances. Because no dimension is specified, CloudWatch returns statistics for all dimensions in the AWS/EC2 namespace.

**Important**
This technique for retrieving all dimensions across an AWS namespace does not work for custom namespaces that you publish to Amazon CloudWatch. With custom namespaces, you must specify the complete set of dimensions that are associated with any given data point to retrieve statistics that include the data point.

**To display average CPU utilization across your instances (console)**

2. In the navigation pane, choose Metrics.
3. Choose the EC2 namespace and then choose Across All Instances.
4. Choose the row that contains CPUUtilization, which displays a graph for the metric for all your EC2 instances. To name the graph, choose the pencil icon. To change the time range, select one of the predefined values or choose custom.

5. To change the statistic or the period for the metric, choose the Graphed metrics tab. Choose the column heading or an individual value, and then choose a different value.

**To get average CPU utilization across your instances (AWS CLI)**

Use the get-metric-statistics command as follows to get the average of the CPUUtilization metric across your instances.

```
aws cloudwatch get-metric-statistics --namespace AWS/EC2 --metric-name CPUUtilization --period 3600 --statistics "Average" "SampleCount" --start-time 2016-10-11T23:18:00 --end-time 2016-10-12T23:18:00
```

The following is example output:
Aggregate Statistics by Auto Scaling Group

You can aggregate statistics for the EC2 instances in an Auto Scaling group. Note that Amazon CloudWatch cannot aggregate data across regions. Metrics are completely separate between regions.

This example shows you how to retrieve the total bytes written to disk for one Auto Scaling group. The total is computed for one-minute periods for a 24-hour interval across all EC2 instances in the specified Auto Scaling group.

To display DiskWriteBytes for the instances in an Auto Scaling group (console)

2. In the navigation pane, choose Metrics.
3. Choose the EC2 namespace and then choose By Auto Scaling Group.
4. Choose the row for the DiskWriteBytes metric and the specific Auto Scaling group, which displays a graph for the metric for the instances in the Auto Scaling group. To name the graph, choose the pencil icon. To change the time range, select one of the predefined values or choose custom.
5. To change the statistic or the period for the metric, choose the Graphed metrics tab. Choose the column heading or an individual value, and then choose a different value.

To display DiskWriteBytes for the instances in an Auto Scaling group (AWS CLI)

Use the get-metric-statistics command as follows.

```bash
aws cloudwatch get-metric-statistics --namespace AWS/EC2 --metric-name DiskWriteBytes --period 360 \n--statistics "Sum" "SampleCount" --dimensions Name=AutoScalingGroupName,Value=my-asg --start-time 2016-10-16T23:18:00 --end-time 2016-10-18T23:18:00
```

The following is example output:

```
{"Datapoints": [ 
{ "SampleCount": 238.0,  
"Timestamp": "2016-10-12T07:18:00Z",  
"Average": 0.038235294117647062,  
"Unit": "Percent" 
}, 
{ "SampleCount": 240.0,  
"Timestamp": "2016-10-12T09:18:00Z",  
"Average": 0.16670833333333332,  
"Unit": "Percent" 
}, 
{ "SampleCount": 238.0,  
"Timestamp": "2016-10-11T23:18:00Z",  
"Average": 0.041596638655462197,  
"Unit": "Percent" 
}, 
... 
"Label": "CPUUtilization" }
```
Aggregate Statistics by AMI

You can aggregate statistics for your instances that have detailed monitoring enabled. Instances that use basic monitoring are not included. Note that Amazon CloudWatch cannot aggregate data across regions. Metrics are completely separate between regions.

Before you can get statistics aggregated across instances, you must enable detailed monitoring (at an additional charge), which provides data in 1-minute periods. For more information, see Enable or Disable Detailed Monitoring for Your Instances (p. 546).

This example shows you how to determine average CPU utilization for all instances that use a specific Amazon Machine Image (AMI). The average is over 60-second time intervals for a one-day period.

To display the average CPU utilization by AMI (console)

2. In the navigation pane, choose Metrics.
3. Choose the EC2 namespace and then choose By Image (AMI) Id.
4. Choose the row for the CPUUtilization metric and the specific AMI, which displays a graph for the metric for the specified AMI. To name the graph, choose the pencil icon. To change the time range, select one of the predefined values or choose custom.
5. To change the statistic or the period for the metric, choose the Graphed metrics tab. Choose the column heading or an individual value, and then choose a different value.

To get the average CPU utilization for an image ID (AWS CLI)

Use the get-metric-statistics command as follows.

```bash
aws cloudwatch get-metric-statistics --namespace AWS/EC2 --metric-name CPUUtilization -- period 3600 \ --statistics Average --dimensions Name=ImageId,Value=ami-3c47a355 --start-time 2016-10-10T00:00:00 --end-time 2016-10-11T00:00:00
```

The following is example output. Each value represents an average CPU utilization percentage for the EC2 instances running the specified AMI.

```json
{
   "Datapoints": [
   {
      "Timestamp": "2016-10-10T07:00:00Z",
      "Average": 0.04100000000000009,
      "Unit": "Percent"
   },
   {
```
Graph Metrics for Your Instances

After you launch an instance, you can open the Amazon EC2 console and view the monitoring graphs for an instance on the Monitoring tab. Each graph is based on one of the available Amazon EC2 metrics.

The following graphs are available:

- Average CPU Utilization (Percent)
- Average Disk Reads (Bytes)
- Average Disk Writes (Bytes)
- Maximum Network In (Bytes)
- Maximum Network Out (Bytes)
- Summary Disk Read Operations (Count)
- Summary Disk Write Operations (Count)
- Summary Status (Any)
- Summary Status Instance (Count)
- Summary Status System (Count)

For more information about the metrics and the data they provide to the graphs, see List the Available CloudWatch Metrics for Your Instances (p. 548).

Graph Metrics Using the CloudWatch Console

You can also use the CloudWatch console to graph metric data generated by Amazon EC2 and other AWS services. For more information, see Graph Metrics in the Amazon CloudWatch User Guide.

Create a CloudWatch Alarm for an Instance

You can create a CloudWatch alarm that monitors CloudWatch metrics for one of your instances. CloudWatch will automatically send you a notification when the metric reaches a threshold you specify. You can create a CloudWatch alarm using the Amazon EC2 console, or using the more advanced options provided by the CloudWatch console.

To create an alarm using the CloudWatch console

For examples, see Creating Amazon CloudWatch Alarms in the Amazon CloudWatch User Guide.

To create an alarm using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance.
4. On the Monitoring tab, choose Create Alarm.

5. In the Create Alarm dialog box, do the following:

   a. Choose create topic. For Send a notification to, enter a name for the SNS topic. For With these recipients, enter one or more email addresses to receive notification.

   b. Specify the metric and the criteria for the policy. For example, you can leave the default settings for Whenever (Average of CPU Utilization). For Is, choose >= and enter 80 percent. For For at least, enter 1 consecutive period of 5 Minutes.

   c. Choose Create Alarm.

Create Alarms That Stop, Terminate, Reboot, or Recover an Instance

Using Amazon CloudWatch alarm actions, you can create alarms that automatically stop, terminate, reboot, or recover your instances. You can use the stop or terminate actions to help you save money when you no longer need an instance to be running. You can use the reboot and recover actions to automatically reboot those instances or recover them onto new hardware if a system impairment occurs.

The AWSServiceRoleForCloudWatchEvents service-linked role enables AWS to perform alarm actions on your behalf. The first time you create an alarm in the AWS Management Console, the IAM CLI, or the IAM API, CloudWatch creates the service-linked role for you.

There are a number of scenarios in which you might want to automatically stop or terminate your instance. For example, you might have instances dedicated to batch payroll processing jobs or scientific computing tasks that run for a period of time and then complete their work. Rather than letting those instances sit idle (and accrue charges), you can stop or terminate them, which can help you to save money. The main difference between using the stop and the terminate alarm actions is that you can easily restart a stopped instance if you need to run it again later, and you can keep the same instance ID and root volume. However, you cannot restart a terminated instance. Instead, you must launch a new instance.
You can add the stop, terminate, reboot, or recover actions to any alarm that is set on an Amazon EC2 per-instance metric, including basic and detailed monitoring metrics provided by Amazon CloudWatch (in the AWS/EC2 namespace), as well as any custom metrics that include the InstanceId dimension, as long as its value refers to a valid running Amazon EC2 instance.

**Console Support**

You can create alarms using the Amazon EC2 console or the CloudWatch console. The procedures in this documentation use the Amazon EC2 console. For procedures that use the CloudWatch console, see Create Alarms That Stop, Terminate, Reboot, or Recover an Instance in the Amazon CloudWatch User Guide.

**Permissions**

If you are an AWS Identity and Access Management (IAM) user, you must have the following permissions to create or modify an alarm:

- `iam:CreateServiceLinkedRole`, `iam:GetPolicy`, `iam:GetPolicyVersion`, and `iam:GetRole` – For all alarms with Amazon EC2 actions
- `ec2:DescribeInstanceStatus` and `ec2:DescribeInstances` – For all alarms on Amazon EC2 instance status metrics
- `ec2:StopInstances` – For alarms with stop actions
- `ec2:TerminateInstances` – For alarms with terminate actions
- No specific permissions are needed for alarms with recover actions.

If you have read/write permissions for Amazon CloudWatch but not for Amazon EC2, you can still create an alarm but the stop or terminate actions won’t be performed on the Amazon EC2 instance. However, if you are later granted permission to use the associated Amazon EC2 APIs, the alarm actions you created earlier are performed. For more information about IAM permissions, see Permissions and Policies in the IAM User Guide.

**Contents**

- Adding Stop Actions to Amazon CloudWatch Alarms (p. 566)
- Adding Terminate Actions to Amazon CloudWatch Alarms (p. 567)
- Adding Reboot Actions to Amazon CloudWatch Alarms (p. 568)
- Adding Recover Actions to Amazon CloudWatch Alarms (p. 569)
- Using the Amazon CloudWatch Console to View Alarm and Action History (p. 570)
- Amazon CloudWatch Alarm Action Scenarios (p. 570)

**Adding Stop Actions to Amazon CloudWatch Alarms**

You can create an alarm that stops an Amazon EC2 instance when a certain threshold has been met. For example, you may run development or test instances and occasionally forget to shut them off. You can create an alarm that is triggered when the average CPU utilization percentage has been lower than 10 percent for 24 hours, signaling that it is idle and no longer in use. You can adjust the threshold, duration, and period to suit your needs, plus you can add an Amazon Simple Notification Service (Amazon SNS) notification so that you receive an email when the alarm is triggered.

Instances that use an Amazon EBS volume as the root device can be stopped or terminated, whereas instances that use the instance store as the root device can only be terminated.

**To create an alarm to stop an idle instance (Amazon EC2 console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.

4. In the Create Alarm dialog box, do the following:

   a. To receive an email when the alarm is triggered, for Send a notification to, choose an existing Amazon SNS topic, or choose create topic to create a new one.

      To create a new topic, for Send a notification to, enter a name for the topic, and then for With these recipients, enter the email addresses of the recipients (separated by commas). After you create the alarm, you will receive a subscription confirmation email that you must accept before you can get notifications for this topic.

   b. Choose Take the action, Stop this instance.

   c. For Whenever, choose the statistic you want to use and then choose the metric. In this example, choose Average and CPU Utilization.

   d. For Is, specify the metric threshold. In this example, enter 10 percent.

   e. For For at least, specify the evaluation period for the alarm. In this example, enter 24 consecutive period(s) of 1 Hour.

   f. To change the name of the alarm, for Name of alarm, enter a new name. Alarm names must contain only ASCII characters.

      If you don’t enter a name for the alarm, Amazon CloudWatch automatically creates one for you.

      **Note**

      You can adjust the alarm configuration based on your own requirements before creating the alarm, or you can edit them later. This includes the metric, threshold, duration, action, and notification settings. However, after you create an alarm, you cannot edit its name later.

   g. Choose Create Alarm.

### Adding Terminate Actions to Amazon CloudWatch Alarms

You can create an alarm that terminates an EC2 instance automatically when a certain threshold has been met (as long as termination protection is not enabled for the instance). For example, you might want to terminate an instance when it has completed its work, and you don’t need the instance again. If you might want to use the instance later, you should stop the instance instead of terminating it. For information on enabling and disabling termination protection for an instance, see Enabling Termination Protection for an Instance in the Amazon EC2 User Guide for Linux Instances.

**To create an alarm to terminate an idle instance (Amazon EC2 console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.

2. In the navigation pane, choose Instances.


4. In the Create Alarm dialog box, do the following:

   a. To receive an email when the alarm is triggered, for Send a notification to, choose an existing Amazon SNS topic, or choose create topic to create a new one.

      To create a new topic, for Send a notification to, enter a name for the topic, and then for With these recipients, enter the email addresses of the recipients (separated by commas). After you create the alarm, you will receive a subscription confirmation email that you must accept before you can get notifications for this topic.

   b. Choose Take the action, Terminate this instance.

   c. For Whenever, choose a statistic and then choose the metric. In this example, choose Average and CPU Utilization.
Create Alarms That Stop, Terminate, Reboot, or Recover an Instance

d. For Is, specify the metric threshold. In this example, enter 10 percent.
e. For For at least, specify the evaluation period for the alarm. In this example, enter 24 consecutive period(s) of 1 Hour.
f. To change the name of the alarm, for Name of alarm, enter a new name. Alarm names must contain only ASCII characters.

If you don't enter a name for the alarm, Amazon CloudWatch automatically creates one for you.

Note
You can adjust the alarm configuration based on your own requirements before creating the alarm, or you can edit them later. This includes the metric, threshold, duration, action, and notification settings. However, after you create an alarm, you cannot edit its name later.
g. Choose Create Alarm.

Adding Reboot Actions to Amazon CloudWatch Alarms

You can create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance and automatically reboots the instance. The reboot alarm action is recommended for Instance Health Check failures (as opposed to the recover alarm action, which is suited for System Health Check failures). An instance reboot is equivalent to an operating system reboot. In most cases, it takes only a few minutes to reboot your instance. When you reboot an instance, it remains on the same physical host, so your instance keeps its public DNS name, private IP address, and any data on its instance store volumes.

Rebooting an instance doesn't start a new instance billing hour, unlike stopping and restarting your instance. For more information, see Reboot Your Instance in the Amazon EC2 User Guide for Linux Instances.

Important
To avoid a race condition between the reboot and recover actions, avoid setting the same number of evaluation periods for a reboot alarm and a recover alarm. We recommend that you set reboot alarms to three evaluation periods of one minute each. For more information, see Evaluating an Alarm in the Amazon CloudWatch User Guide.

To create an alarm to reboot an instance (Amazon EC2 console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
4. In the Create Alarm dialog box, do the following:
   a. To receive an email when the alarm is triggered, for Send a notification to, choose an existing Amazon SNS topic, or choose create topic to create a new one.
      To create a new topic, for Send a notification to, enter a name for the topic, and for With these recipients, enter the email addresses of the recipients (separated by commas). After you create the alarm, you will receive a subscription confirmation email that you must accept before you can get notifications for this topic.
   b. Select Take the action, Reboot this instance.
   c. For Whenever, choose Status Check Failed (Instance).
   d. For For at least, specify the evaluation period for the alarm. In this example, enter 3 consecutive period(s) of 1 Minute.
   e. To change the name of the alarm, for Name of alarm, enter a new name. Alarm names must contain only ASCII characters.
      If you don't enter a name for the alarm, Amazon CloudWatch automatically creates one for you.
Adding Recover Actions to Amazon CloudWatch Alarms

You can create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance. If the instance becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair, you can automatically recover the instance. Terminated instances cannot be recovered. A recovered instance is identical to the original instance, including the instance ID, private IP addresses, Elastic IP addresses, and all instance metadata.

CloudWatch prevents you from adding a recovery action to an alarm that is on an instance which does not support recovery actions.

When the StatusCheckFailed_System alarm is triggered, and the recover action is initiated, you are notified by the Amazon SNS topic that you chose when you created the alarm and associated the recover action. During instance recovery, the instance is migrated during an instance reboot, and any data that is in-memory is lost. When the process is complete, information is published to the SNS topic you’ve configured for the alarm. Anyone who is subscribed to this SNS topic receives an email notification that includes the status of the recovery attempt and any further instructions. You notice an instance reboot on the recovered instance.

The recover action can be used only with StatusCheckFailed_System, not with StatusCheckFailed_Instance.

The following problems can cause system status checks to fail:

- Loss of network connectivity
- Loss of system power
- Software issues on the physical host
- Hardware issues on the physical host that impact network reachability

The recover action is supported only on instances with the following characteristics:

- Use one of the following instance types: C3, C4, C5, C5n, M3, M4, M5, M5a, P3, R3, R4, R5, R5a, T2, T3, T3a, X1, or X1e
- Use default or dedicated instance tenancy
- Use EBS volumes only (do not configure instance store volumes). For more information, see ‘Recover this instance’ is disabled.

If your instance has a public IP address, it retains the public IP address after recovery.

**Important**

To avoid a race condition between the reboot and recover actions, avoid setting the same number of evaluation periods for a reboot alarm and a recover alarm. We recommend that you set recover alarms to two evaluation periods of one minute each. For more information, see Evaluating an Alarm in the Amazon CloudWatch User Guide.

**To create an alarm to recover an instance (Amazon EC2 console)**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**.
3. Select the instance. On the **Monitoring** tab, choose **Create Alarm**.
4. In the **Create Alarm** dialog box, do the following:
a. To receive an email when the alarm is triggered, for **Send a notification to**, choose an existing Amazon SNS topic, or choose **create topic** to create a new one.

To create a new topic, for **Send a notification to**, enter a name for the topic, and for **With these recipients**, enter the email addresses of the recipients (separated by commas). After you create the alarm, you will receive a subscription confirmation email that you must accept before you can get email for this topic.

**Note**

- Users must subscribe to the specified SNS topic to receive email notifications when the alarm is triggered.
- The AWS account root user always receives email notifications when automatic instance recovery actions occur, even if an SNS topic is not specified.
- The AWS account root user always receives email notifications when automatic instance recovery actions occur, even if it is not subscribed to the specified SNS topic.

b. Select **Take the action**, **Recover this instance**.

c. For **Whenever**, choose **Status Check Failed (System)**.

d. For **For at least**, specify the evaluation period for the alarm. In this example, enter 2 consecutive period(s) of **1 Minute**.

e. To change the name of the alarm, for **Name of alarm**, enter a new name. Alarm names must contain only ASCII characters.

    If you don’t enter a name for the alarm, Amazon CloudWatch automatically creates one for you.

f. Choose **Create Alarm**.

### Using the Amazon CloudWatch Console to View Alarm and Action History

You can view alarm and action history in the Amazon CloudWatch console. Amazon CloudWatch keeps the last two weeks’ worth of alarm and action history.

**To view the history of triggered alarms and actions (CloudWatch console)**

2. In the navigation pane, choose **Alarms**.
3. Select an alarm.
4. The **Details** tab shows the most recent state transition along with the time and metric values.
5. Choose the **History** tab to view the most recent history entries.

### Amazon CloudWatch Alarm Action Scenarios

You can use the Amazon EC2 console to create alarm actions that stop or terminate an Amazon EC2 instance when certain conditions are met. In the following screen capture of the console page where you set the alarm actions, we’ve numbered the settings. We’ve also numbered the settings in the scenarios that follow, to help you create the appropriate actions.
Scenario 1: Stop Idle Development and Test Instances

Create an alarm that stops an instance used for software development or testing when it has been idle for at least an hour.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stop</td>
</tr>
<tr>
<td>2</td>
<td>Maximum</td>
</tr>
<tr>
<td>3</td>
<td>CPUUtilization</td>
</tr>
<tr>
<td>4</td>
<td>&lt;=</td>
</tr>
<tr>
<td>5</td>
<td>10%</td>
</tr>
<tr>
<td>6</td>
<td>60 minutes</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Scenario 2: Stop Idle Instances

Create an alarm that stops an instance and sends an email when the instance has been idle for 24 hours.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stop and email</td>
</tr>
<tr>
<td>2</td>
<td>Average</td>
</tr>
</tbody>
</table>
### Create Alarms That Stop, Terminate, Reboot, or Recover an Instance

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>CPUUtilization</td>
</tr>
<tr>
<td>4</td>
<td>&lt;=</td>
</tr>
<tr>
<td>5</td>
<td>5%</td>
</tr>
<tr>
<td>6</td>
<td>60 minutes</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
</tr>
</tbody>
</table>

#### Scenario 3: Send Email About Web Servers with Unusually High Traffic

Create an alarm that sends email when an instance exceeds 10 GB of outbound network traffic per day.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Email</td>
</tr>
<tr>
<td>2</td>
<td>Sum</td>
</tr>
<tr>
<td>3</td>
<td>NetworkOut</td>
</tr>
<tr>
<td>4</td>
<td>&gt;</td>
</tr>
<tr>
<td>5</td>
<td>10 GB</td>
</tr>
<tr>
<td>6</td>
<td>1 day</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

#### Scenario 4: Stop Web Servers with Unusually High Traffic

Create an alarm that stops an instance and send a text message (SMS) if outbound traffic exceeds 1 GB per hour.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stop and send SMS</td>
</tr>
<tr>
<td>2</td>
<td>Sum</td>
</tr>
<tr>
<td>3</td>
<td>NetworkOut</td>
</tr>
<tr>
<td>4</td>
<td>&gt;</td>
</tr>
<tr>
<td>5</td>
<td>1 GB</td>
</tr>
<tr>
<td>6</td>
<td>1 hour</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

#### Scenario 5: Stop an Instance Experiencing a Memory Leak

Create an alarm that stops an instance when memory utilization reaches or exceeds 90%, so that application logs can be retrieved for troubleshooting.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Note
The MemoryUtilization metric is a custom metric. In order to use the MemoryUtilization metric, you must install the Perl scripts for Linux instances. For more information, see Monitoring Memory and Disk Metrics for Amazon EC2 Linux Instances.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stop</td>
</tr>
<tr>
<td>2</td>
<td>Maximum</td>
</tr>
<tr>
<td>3</td>
<td>MemoryUtilization</td>
</tr>
<tr>
<td>4</td>
<td>&gt;=</td>
</tr>
<tr>
<td>5</td>
<td>90%</td>
</tr>
<tr>
<td>6</td>
<td>1 minute</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Scenario 6: Stop an Impaired Instance
Create an alarm that stops an instance that fails three consecutive status checks (performed at 5-minute intervals).

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stop</td>
</tr>
<tr>
<td>2</td>
<td>Average</td>
</tr>
<tr>
<td>3</td>
<td>StatusCheckFailed_System</td>
</tr>
<tr>
<td>4</td>
<td>&gt;=</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>15 minutes</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Scenario 7: Terminate Instances When Batch Processing Jobs Are Complete
Create an alarm that terminates an instance that runs batch jobs when it is no longer sending results data.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Terminate</td>
</tr>
<tr>
<td>2</td>
<td>Maximum</td>
</tr>
<tr>
<td>3</td>
<td>NetworkOut</td>
</tr>
<tr>
<td>4</td>
<td>&lt;=</td>
</tr>
</tbody>
</table>
Automating Amazon EC2 with CloudWatch Events

Amazon CloudWatch Events enables you to automate your AWS services and respond automatically to system events such as application availability issues or resource changes. Events from AWS services are delivered to CloudWatch Events in near real time. You can write simple rules to indicate which events are of interest to you, and the automated actions to take when an event matches a rule. The actions that can be automatically triggered include the following:

- Invoking an AWS Lambda function
- Invoking Amazon EC2 Run Command
- Relaying the event to Amazon Kinesis Data Streams
- Activating an AWS Step Functions state machine
- Notifying an Amazon SNS topic or an Amazon SQS queue

Some examples of using CloudWatch Events with Amazon EC2 include:

- Activating a Lambda function whenever a new Amazon EC2 instance starts.
- Notifying an Amazon SNS topic when an Amazon EBS volume is created or modified.
- Sending a command to one or more Amazon EC2 instances using Amazon EC2 Run Command whenever a certain event in another AWS service occurs.

For more information, see the Amazon CloudWatch Events User Guide.

Logging Amazon EC2 and Amazon EBS API Calls with AWS CloudTrail

Amazon EC2 and Amazon EBS are integrated with AWS CloudTrail, a service that provides a record of actions taken by a user, role, or an AWS service in Amazon EC2 and Amazon EBS. CloudTrail captures all API calls for Amazon EC2 and Amazon EBS as events, including calls from the console and from code calls to the APIs. If you create a trail, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Amazon EC2 and Amazon EBS. If you don't configure a trail, you can still view the most recent events in the CloudTrail console in Event history. Using the information collected by CloudTrail, you can determine the request that was made to Amazon EC2 and Amazon EBS, the IP address from which the request was made, who made the request, when it was made, and additional details.

To learn more about CloudTrail, see the AWS CloudTrail User Guide.
Amazon EC2 and Amazon EBS Information in CloudTrail

CloudTrail is enabled on your AWS account when you create the account. When activity occurs in Amazon EC2 and Amazon EBS, that activity is recorded in a CloudTrail event along with other AWS service events in Event history. You can view, search, and download recent events in your AWS account. For more information, see Viewing Events with CloudTrail Event History.

For an ongoing record of events in your AWS account, including events for Amazon EC2 and Amazon EBS, create a trail. A trail enables CloudTrail to deliver log files to an Amazon S3 bucket. By default, when you create a trail in the console, the trail applies to all Regions. The trail logs events from all Regions in the AWS partition and delivers the log files to the Amazon S3 bucket that you specify. Additionally, you can configure other AWS services to further analyze and act upon the event data collected in CloudTrail logs. For more information, see:

- Overview for Creating a Trail
- CloudTrail Supported Services and Integrations
- Configuring Amazon SNS Notifications for CloudTrail
- Receiving CloudTrail Log Files from Multiple Regions and Receiving CloudTrail Log Files from Multiple Accounts

All Amazon EC2 and Amazon EBS actions are logged by CloudTrail and are documented in the Amazon EC2 API Reference. For example, calls to the RunInstances, DescribeInstances, or CreateImage actions generate entries in the CloudTrail log files.

Every event or log entry contains information about who generated the request. The identity information helps you determine the following:

- Whether the request was made with root or IAM user credentials.
- Whether the request was made with temporary security credentials for a role or federated user.
- Whether the request was made by another AWS service.

For more information, see the CloudTrail userIdentity Element.

Understanding Amazon EC2 and Amazon EBS Log File Entries

A trail is a configuration that enables delivery of events as log files to an Amazon S3 bucket that you specify. CloudTrail log files contain one or more log entries. An event represents a single request from any source and includes information about the requested action, the date and time of the action, request parameters, and so on. CloudTrail log files are not an ordered stack trace of the public API calls, so they do not appear in any specific order.

The following log file record shows that a user terminated an instance.

```json
{
   "Records":[
      {
         "eventVersion":"1.03",
         "userIdentity":{
            "type":"Root",
            "principalId":"123456789012",
            "arn":"arn:aws:iam::123456789012:root",
```
Using AWS CloudTrail to Audit Users that Connect via EC2 Instance Connect

Use AWS CloudTrail to audit the users that connect to your instances via EC2 Instance Connect.

To audit SSH activity via EC2 Instance Connect using the AWS CloudTrail console

1. Open the AWS CloudTrail console at https://console.aws.amazon.com/cloudtrail/.
2. Verify that you are in the correct Region.
3. In the navigation pane, choose Event history.
5. (Optional) For Time range, select a time range.
6. Choose the Refresh events icon.
7. The page displays the events that correspond to the SendSSHPublicKey API calls. Expand an event using the arrow to view additional details, such as the user name and AWS access key that was used to make the SSH connection, and the source IP address.
8. To display the full event information in JSON format, choose View event. The requestParameters field contains the destination instance ID, OS user name, and public key that were used to make the SSH connection.
If you have configured your AWS account to collect CloudTrail events in an S3 bucket, you can download and audit the information programmatically. For more information, see Getting and Viewing Your CloudTrail Log Files in the AWS CloudTrail User Guide.
Network and Security

Amazon EC2 provides the following network and security features.

Features

- Amazon EC2 Key Pairs and Windows Instances (p. 578)
- Amazon EC2 Security Groups for Windows Instances (p. 583)
- Controlling Access to Amazon EC2 Resources (p. 597)
- Amazon EC2 Instance IP Addressing (p. 656)
- Bring Your Own IP Addresses (BYOIP) (p. 670)
- Elastic IP Addresses (p. 674)
- Elastic Network Interfaces (p. 679)
- Enhanced Networking on Windows (p. 700)
- Placement Groups (p. 710)
- Network Maximum Transmission Unit (MTU) for Your EC2 Instance (p. 718)
- Virtual Private Clouds (p. 721)
- Ports and Protocols for Windows Amazon Machine Images (AMIs) (p. 722)
- EC2-Classic (p. 752)

Amazon EC2 Key Pairs and Windows Instances

Amazon EC2 uses public–key cryptography to encrypt and decrypt login information. Public–key cryptography uses a public key to encrypt a piece of data, and then the recipient uses the private key to decrypt the data. The public and private keys are known as a key pair. Public-key cryptography enables you to securely access your instances using a private key instead of a password.

When you launch an instance, you specify the key pair. You can specify an existing key pair or a new key pair that you create at launch. With Windows instances, you use the private key to obtain the administrator password and then log in using RDP. For more information about key pairs and Linux instances, see Amazon EC2 Key Pairs in the Amazon EC2 User Guide for Linux Instances.

Creating a Key Pair

You can use Amazon EC2 to create your key pair. For more information, see Creating a Key Pair Using Amazon EC2 (p. 579).

Alternatively, you could use a third-party tool and then import the public key to Amazon EC2. For more information, see Importing Your Own Public Key to Amazon EC2 (p. 579).

Each key pair requires a name. Be sure to choose a name that is easy to remember. Amazon EC2 associates the public key with the name that you specify as the key name.

Amazon EC2 stores the public key only, and you store the private key. Anyone who possesses your private key can decrypt your login information, so it's important that you store your private keys in a secure place.

The keys that Amazon EC2 uses are 2048-bit SSH-2 RSA keys. You can have up to five thousand key pairs per Region.

Launching and Connecting to Your Instance
When you launch an instance, you should specify the name of the key pair you plan to use to connect to the instance. If you don't specify the name of an existing key pair when you launch an instance, you won't be able to connect to the instance. When you connect to the instance, you must specify the private key that corresponds to the key pair you specified when you launched the instance.

Contents
- Creating a Key Pair Using Amazon EC2 (p. 579)
- Importing Your Own Public Key to Amazon EC2 (p. 579)
- Retrieving the Public Key for Your Key Pair on Linux (p. 581)
- Retrieving the Public Key for Your Key Pair on Windows (p. 581)
- Retrieving the Public Key for Your Key Pair From Your Instance (p. 581)
- Verifying Your Key Pair's Fingerprint (p. 582)
- Deleting Your Key Pair (p. 582)
- Connecting to Your Windows Instance if You Lose Your Private Key (p. 583)

Creating a Key Pair Using Amazon EC2

You can create a key pair using the Amazon EC2 console or the command line. After you create a key pair, you can specify it when you launch your instance.

To create your key pair using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, under NETWORK & SECURITY, choose Key Pairs.
   Note
   The navigation pane is on the left side of the Amazon EC2 console. If you do not see the pane, it might be minimized; choose the arrow to expand the pane.
3. Choose Create Key Pair.
4. For Key pair name, enter a name for the new key pair, and then choose Create.
5. The private key file is automatically downloaded by your browser. The base file name is the name you specified as the name of your key pair, and the file name extension is .pem. Save the private key file in a safe place.
   Important
   This is the only chance for you to save the private key file. You'll need to provide the name of your key pair when you launch an instance and the corresponding private key each time you connect to the instance.

To create your key pair using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
- create-key-pair (AWS CLI)
- New-EC2KeyPair (AWS Tools for Windows PowerShell)

Importing Your Own Public Key to Amazon EC2

Instead of using Amazon EC2 to create your key pair, you can create an RSA key pair using a third-party tool and then import the public key to Amazon EC2. For example, you can use ssh-keygen (a tool provided with the standard OpenSSH installation) to create a key pair. Alternatively, Java, Ruby, Python,
and many other programming languages provide standard libraries that you can use to create an RSA key pair.

Requirements

- The following formats are supported:
  - OpenSSH public key format
  - Base64 encoded DER format
  - SSH public key file format as specified in RFC4716
  - SSH private key file format must be PEM (for example, use `ssh-keygen -m PEM` to convert the OpenSSH key into the PEM format)
- Create an RSA key. Amazon EC2 does not accept DSA keys.
- The supported lengths are 1024, 2048, and 4096.

To create a key pair using a third-party tool

1. Generate a key pair with a third-party tool of your choice.
2. Save the public key to a local file. For example, `C:\keys\my-key-pair.pub`. The file name extension for this file is not important.
3. Save the private key to a different local file that has the `.pem` extension. For example, `C:\keys\my-key-pair.pem`. Save the private key file in a safe place. You'll need to provide the name of your key pair when you launch an instance and the corresponding private key each time you connect to the instance.

Use the following steps to import your key pair using the Amazon EC2 console.

To import the public key

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, under NETWORK & SECURITY, choose Key Pairs.
3. Choose Import Key Pair.
4. In the Import Key Pair dialog box, choose Browse, and select the public key file that you saved previously. Enter a name for the key pair in the Key pair name field, and choose Import.

To import the public key using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `import-key-pair` (AWS CLI)
- `Import-EC2KeyPair` (AWS Tools for Windows PowerShell)

After the public key file is imported, you can verify that the key pair was imported successfully using the Amazon EC2 console as follows.

To verify that your key pair was imported

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region in which you created the key pair.
3. In the navigation pane, under NETWORK & SECURITY, choose Key Pairs.
4. Verify that the key pair that you imported is in the displayed list of key pairs.
To view your key pair using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `describe-key-pairs` (AWS CLI)
- `Get-EC2KeyPair` (AWS Tools for Windows PowerShell)

Retrieving the Public Key for Your Key Pair on Linux

On your local Linux or Mac computer, you can use the `ssh-keygen` command to retrieve the public key for your key pair. Specify the path where you downloaded your private key (the `.pem` file).

```
ssh-keygen -y -f /path_to_key_pair/my-key-pair.pem
```

The command returns the public key. For example:

```
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAABAQClKsfkNkuSevGj3eYhCe53pcjqp3maAhDFcvBS7O6V
hs2ItxcIh+PnDSUaw+WNQn/m2phTk/a/gU8jZeoOWbkM4yxyyb/wB96xbiFveSFJuo/p/d6RJhJ0I0iBxr
lsJnBagntcklj7Fbt+xMXLvWvJryDUiiBMFjtYtWB+QhYXUMOzce5Pjz5/186e7jvnV3iAoG/cQk+0FzZ
qaeJAAHco+CY/5WkUBkrmFjx6HcXkvJ3MPkYQSxqC0+FmUZofe221Ct5IMUXxXPrX4rWi+e7wB3Rb
BoqOq2d8v7yeb7OzlnWoyN0qFU0X2A46RA8qPYiCNYwI3f05p6KLxEXAMPLE
```

If the command fails, ensure that you've changed the permissions on your key pair file so that only you can view it by running the following command:

```
chmod 400 my-key-pair.pem
```

Retrieving the Public Key for Your Key Pair on Windows

On your local Windows computer, you can use PuTTYgen to get the public key for your key pair.

Start PuTTYgen and choose `Load`. Select the `.ppk` or `.pem` file. PuTTYgen displays the public key under `Public key for pasting into OpenSSH authorized_keys file`. You can also view the public key by choosing `Save public key`, specifying a name for the file, saving the file, and then opening the file.

Retrieving the Public Key for Your Key Pair From Your Instance

The public key that you specified when you launched an instance is also available to you through its instance metadata. To view the public key that you specified when launching the instance, use the following command from your instance:

```
```

```
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAABAQClKsfkNkuSevGj3eYhCe53pcjqp3maAhDFcvBS7O6V
hs2ItxcIh+PnDSUaw+WNQn/m2phTk/a/gU8jZeoOWbkM4yxyyb/wB96xbiFveSFJuo/p/d6RJhJ0I0iBxr
lsJnBagntcklj7Fbt+xMXLvWvJryDUiiBMFjtYtWB+QhYXUMOzce5Pjz5/186e7jvnV3iAoG/cQk+0FzZ
```
Verifying Your Key Pair’s Fingerprint

On the **Key Pairs** page in the Amazon EC2 console, the **Fingerprint** column displays the fingerprints generated from your key pairs. AWS calculates the fingerprint differently depending on whether the key pair was generated by AWS or a third-party tool. If you created the key pair using AWS, the fingerprint is calculated using an SHA-1 hash function. If you created the key pair with a third-party tool and uploaded the public key to AWS, or if you generated a new public key from an existing AWS-created private key and uploaded it to AWS, the fingerprint is calculated using an MD5 hash function.

You can use the SSH2 fingerprint that's displayed on the **Key Pairs** page to verify that the private key you have on your local machine matches the public key stored in AWS. From the computer where you downloaded the private key file, generate an SSH2 fingerprint from the private key file. The output should match the fingerprint that's displayed in the console.

If you created your key pair using AWS, you can use the OpenSSL tools to generate a fingerprint as follows:

```bash
C:\> openssl pkcs8 -in path_to_private_key -inform PEM -outform DER -topk8 -nocrypt | openssl sha1 -c
```

If you created a key pair using a third-party tool and uploaded the public key to AWS, you can use the OpenSSL tools to generate the fingerprint as follows:

```bash
C:\> openssl rsa -in path_to_private_key -pubout -outform DER | openssl md5 -c
```

If you created an OpenSSH key pair using OpenSSH 7.8 or later and uploaded the public key to AWS, you can use `ssh-keygen` to generate the fingerprint as follows:

```bash
C:\> ssh-keygen -ef path_to_private_key -m PEM | openssl rsa -RSAPublicKey_in -outform DER | openssl md5 -c
```

Deleting Your Key Pair

When you delete a key pair, you are only deleting Amazon EC2’s copy of the public key. Deleting a key pair doesn’t affect the private key on your computer or the public key on any instances already launched using that key pair. You can’t launch a new instance using a deleted key pair, but you can continue to connect to any instances that you launched using a deleted key pair, as long as you still have the private key (.pem) file.

**Note**

If you’re using an Auto Scaling group (for example, in an Elastic Beanstalk environment), ensure that the key pair you’re deleting is not specified in your launch configuration. Amazon EC2 Auto Scaling launches a replacement instance if it detects an unhealthy instance; however, the instance launch fails if the key pair cannot be found.

You can delete a key pair using the Amazon EC2 console or the command line.
Connecting to Your Windows Instance if You Lose Your Private Key

When you connect to a newly launched Windows instance, you decrypt the password for the Administrator account using the private key for the key pair that you specified when you launched the instance.

If you lose the Administrator password and you no longer have the private key, you must reset the password or create a new instance. For more information, see Resetting a Lost or Expired Windows Administrator Password (p. 1058). For steps to reset the password using an AWS Systems Manager document, see Reset Passwords and SSH Keys on Amazon EC2 Instances in the AWS Systems Manager User Guide.

Amazon EC2 Security Groups for Windows Instances

A security group acts as a virtual firewall that controls the traffic for one or more instances. When you launch an instance, you can specify one or more security groups; otherwise, we use the default security group. You can add rules to each security group that allow traffic to or from its associated instances. You can modify the rules for a security group at any time; the new rules are automatically applied to all instances that are associated with the security group. When we decide whether to allow traffic to reach an instance, we evaluate all the rules from all the security groups that are associated with the instance.

When you launch an instance in a VPC, you must specify a security group that's created for that VPC. After you launch an instance, you can change its security groups. Security groups are associated with network interfaces. Changing an instance's security groups changes the security groups associated with the primary network interface (eth0). For more information, see Changing an Instance's Security Groups in the Amazon VPC User Guide. You can also change the security groups associated with any other network interface. For more information, see Changing the Security Group (p. 695).

If you have requirements that aren't met by security groups, you can maintain your own firewall on any of your instances in addition to using security groups.

If you need to allow traffic to a Linux instance, see Amazon EC2 Security Groups for Linux Instances in the Amazon EC2 User Guide for Linux Instances.
Security Group Rules

The rules of a security group control the inbound traffic that's allowed to reach the instances that are associated with the security group and the outbound traffic that's allowed to leave them.

The following are the characteristics of security group rules:

- By default, security groups allow all outbound traffic.
- Security group rules are always permissive; you can't create rules that deny access.
- Security groups are stateful — if you send a request from your instance, the response traffic for that request is allowed to flow in regardless of inbound security group rules. For VPC security groups, this also means that responses to allowed inbound traffic are allowed to flow out, regardless of outbound rules. For more information, see Connection Tracking (p. 585).
- You can add and remove rules at any time. Your changes are automatically applied to the instances associated with the security group.

  **Note**
  The effect of some rule changes may depend on how the traffic is tracked. For more information, see Connection Tracking (p. 585).

- When you associate multiple security groups with an instance, the rules from each security group are effectively aggregated to create one set of rules. We use this set of rules to determine whether to allow access.

  **Note**
  You can assign multiple security groups to an instance, therefore an instance can have hundreds of rules that apply. This might cause problems when you access the instance. We recommend that you condense your rules as much as possible.
For each rule, you specify the following:

- **Protocol**: The protocol to allow. The most common protocols are 6 (TCP) 17 (UDP), and 1 (ICMP).
- **Port range**: For TCP, UDP, or a custom protocol, the range of ports to allow. You can specify a single port number (for example, 22), or range of port numbers (for example, 7000–8000).
- **ICMP type and code**: For ICMP, the ICMP type and code.
- **Source or destination**: The source (inbound rules) or destination (outbound rules) for the traffic. Specify one of these options:
  - An individual IPv4 address. You must use the /32 prefix length; for example, 203.0.113.1/32.
  - An individual IPv6 address. You must use the /128 prefix length; for example, 2001:db8:1234:1a00::123/128.
  - A range of IPv4 addresses, in CIDR block notation, for example, 203.0.113.0/24.
  - A range of IPv6 addresses, in CIDR block notation, for example, 2001:db8:1234:1a00::/64.
  - The prefix list ID for the AWS service; for example, pl-1a2b3c4d. For more information, see Gateway VPC Endpoints in the Amazon VPC User Guide.
  - Another security group. This allows instances associated with the specified security group to access instances associated with this security group. This does not add rules from the source security group to this security group. You can specify one of the following security groups:
    - The current security group
    - A different security group for the same VPC
    - A different security group for a peer VPC in a VPC peering connection
- **(Optional) Description**: You can add a description for the rule; for example, to help you identify it later. A description can be up to 255 characters in length. Allowed characters are a-z, A-Z, 0-9, spaces, and _.-/\#,@\[]+=;{}!$*.

When you specify a security group as the source or destination for a rule, the rule affects all instances associated with the security group. Incoming traffic is allowed based on the private IP addresses of the instances that are associated with the source security group (and not the public IP or Elastic IP addresses). For more information about IP addresses, see Amazon EC2 Instance IP Addressing (p. 656). If your security group rule references a security group in a peer VPC, and the referenced security group or VPC peering connection is deleted, the rule is marked as stale. For more information, see Working with Stale Security Group Rules in the Amazon VPC Peering Guide.

If there is more than one rule for a specific port, we apply the most permissive rule. For example, if you have a rule that allows access to TCP port 3389 (RDP) from IP address 203.0.113.1 and another rule that allows access to TCP port 3389 from everyone, everyone has access to TCP port 3389.

### Connection Tracking

Your security groups use connection tracking to track information about traffic to and from the instance. Rules are applied based on the connection state of the traffic to determine if the traffic is allowed or denied. This allows security groups to be stateful — responses to inbound traffic are allowed to flow out of the instance regardless of outbound security group rules, and vice versa. For example, if you initiate an ICMP ping command to your instance from your home computer, and your inbound security group rules allow ICMP traffic, information about the connection (including the port information) is tracked. Response traffic from the instance for the ping command is not tracked as a new request, but rather as an established connection and is allowed to flow out of the instance, even if your outbound security group rules restrict outbound ICMP traffic.

Not all flows of traffic are tracked. If a security group rule permits TCP or UDP flows for all traffic (0.0.0.0/0) and there is a corresponding rule in the other direction that permits all response traffic (0.0.0.0/0) for all ports (0-65535), then that flow of traffic is not tracked. The response traffic is therefore allowed to flow based on the inbound or outbound rule that permits the response traffic, and not on tracking information.
In the following example, the security group has specific inbound rules for TCP and ICMP traffic, and an outbound rule that allows all outbound traffic.

<table>
<thead>
<tr>
<th>Inbound rules</th>
<th>Port number</th>
<th>Source IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>22 (SSH)</td>
<td>203.0.113.1/32</td>
</tr>
<tr>
<td>TCP</td>
<td>80 (HTTP)</td>
<td>0.0.0.0/0</td>
</tr>
<tr>
<td>ICMP</td>
<td>All</td>
<td>0.0.0.0/0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outbound rules</th>
<th>Destination IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.0.0.0/0</td>
</tr>
</tbody>
</table>

TCP traffic on port 22 (SSH) to and from the instance is tracked, because the inbound rule allows traffic from 203.0.113.1/32 only, and not all IP addresses (0.0.0.0/0). TCP traffic on port 80 (HTTP) to and from the instance is not tracked, because both the inbound and outbound rules allow all traffic (0.0.0.0/0). ICMP traffic is always tracked, regardless of rules. If you remove the outbound rule from the security group, then all traffic to and from the instance is tracked, including traffic on port 80 (HTTP).

An existing flow of traffic that is tracked may not be interrupted when you remove the security group rule that enables that flow. Instead, the flow is interrupted when it's stopped by you or the other host for at least a few minutes (or up to 5 days for established TCP connections). For UDP, this may require terminating actions on the remote side of the flow. An untracked flow of traffic is immediately interrupted if the rule that enables the flow is removed or modified. For example, if you remove a rule that allows all inbound SSH traffic to the instance, then your existing SSH connections to the instance are immediately dropped.

For protocols other than TCP, UDP, or ICMP, only the IP address and protocol number is tracked. If your instance sends traffic to another host (host B), and host B initiates the same type of traffic to your instance in a separate request within 600 seconds of the original request or response, your instance accepts it regardless of inbound security group rules, because it's regarded as response traffic.

To ensure that traffic is immediately interrupted when you remove a security group rule, or to ensure that all inbound traffic is subject to firewall rules, you can use a network ACL for your subnet — network ACLs are stateless and therefore do not automatically allow response traffic. For more information, see Network ACLs in the Amazon VPC User Guide.

**Default Security Groups**

Your AWS account automatically has a default security group for the default VPC in each region. If you don't specify a security group when you launch an instance, the instance is automatically associated with the default security group for the VPC.

A default security group is named default, and it has an ID assigned by AWS. The following are the default rules for each default security group:

- Allows all inbound traffic from other instances associated with the default security group (the security group specifies itself as a source security group in its inbound rules)
- Allows all outbound traffic from the instance.

You can add or remove inbound and outbound rules for any default security group.
You can't delete a default security group. If you try to delete a default security group, you'll get the following error: `Client.CannotDelete: the specified group: "sg-51530134" name: "default" cannot be deleted by a user.`

Custom Security Groups

If you don't want your instances to use the default security group, you can create your own security groups and specify them when you launch your instances. You can create multiple security groups to reflect the different roles that your instances play; for example, a web server or a database server.

When you create a security group, you must provide it with a name and a description. Security group names and descriptions can be up to 255 characters in length, and are limited to the following characters:

a-z, A-Z, 0-9, spaces, and _-/:()#,@\[]+=&;{}!$*

A security group name cannot start with `sg-`. A security group name must be unique for the VPC.

The following are the default rules for a security group that you create:

- Allows no inbound traffic
- Allows all outbound traffic

After you've created a security group, you can change its inbound rules to reflect the type of inbound traffic that you want to reach the associated instances. You can also change its outbound rules.

For more information about the rules you can add to a security group, see Security Group Rules Reference (p. 591).

Working with Security Groups

You can create, view, update, and delete security groups and security group rules using the Amazon EC2 console.

Tasks

- Creating a Security Group (p. 587)
- Describing Your Security Groups (p. 588)
- Adding Rules to a Security Group (p. 588)
- Updating Security Group Rules (p. 590)
- Deleting Rules from a Security Group (p. 590)
- Deleting a Security Group (p. 591)

Creating a Security Group

You can create a custom security group using the Amazon EC2 console. You must specify the VPC for which you're creating the security group.

To create a new security group using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Security Groups.
4. Specify a name and description for the security group.
5. For VPC, choose the ID of the VPC.
6. You can start adding rules, or you can choose Create to create the security group now (you can always add rules later). For more information about adding rules, see Adding Rules to a Security Group (p. 588).

To create a security group using the command line

- create-security-group (AWS CLI)

The Amazon EC2 console enables you to copy the rules from an existing security group to a new security group.

To copy a security group using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Security Groups.
3. Select the security group you want to copy, choose Actions, Copy to new.
4. The Create Security Group dialog opens, and is populated with the rules from the existing security group. Specify a name and description for your new security group. For VPC, choose the ID of the VPC. When you are done, choose Create.

You can assign a security group to an instance when you launch the instance. When you add or remove rules, those changes are automatically applied to all instances to which you've assigned the security group.

After you launch an instance, you can change its security groups. For more information, see Changing an Instance's Security Groups in the Amazon VPC User Guide.

Describing Your Security Groups

You can view information about your security groups using the Amazon EC2 console or the command line.

To describe your security groups using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Security Groups.
3. (Optional) Select VPC ID from the filter list, then choose the ID of the VPC.
4. Select a security group. We display general information in the Description tab, inbound rules on the Inbound tab, outbound rules on the Outbound tab, and tags on the Tags tab.

To describe one or more security groups using the command line

- describe-security-groups (AWS CLI)

Adding Rules to a Security Group

When you add a rule to a security group, the new rule is automatically applied to any instances associated with the security group after a short period.

For more information about choosing security group rules for specific types of access, see Security Group Rules Reference (p. 591).
To add rules to a security group using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Security Groups and select the security group.
3. On the Inbound tab, choose Edit.
4. In the dialog, choose Add Rule and do the following:
   - For Type, select the protocol.
   - If you select a custom TCP or UDP protocol, specify the port range in Port Range.
   - If you select a custom ICMP protocol, choose the ICMP type name from Protocol, and, if applicable, the code name from Port Range.
   - For Source, choose one of the following:
     - Custom: in the provided field, you must specify an IP address in CIDR notation, a CIDR block, or another security group.
     - Anywhere: adds the 0.0.0.0/0 IPv4 CIDR block. This option enables all traffic of the specified type to reach your instance. This is acceptable for a short time in a test environment, but it's unsafe for production environments. In production, authorize only a specific IP address or range of addresses to access your instance.
     - My IP: adds the IPv4 address of your local computer.
   - For Description, you can optionally specify a description for the rule.

5. Choose Save.
6. You can also specify outbound rules. On the Outbound tab, choose Edit, Add Rule, and do the following:
   - For Type, select the protocol.
   - If you select a custom TCP or UDP protocol, specify the port range in Port Range.
   - If you select a custom ICMP protocol, choose the ICMP type name from Protocol, and, if applicable, the code name from Port Range.
   - For Destination, choose one of the following:
     - Custom: in the provided field, you must specify an IP address in CIDR notation, a CIDR block, or another security group.
     - Anywhere: adds the 0.0.0.0/0 IPv4 CIDR block. This option enables outbound traffic to all IP addresses.
     - My IP: adds the IPv4 address of your local computer.
   - For Description, you can optionally specify a description for the rule.

7. Choose Save.

To add one or more ingress rules to a security group using the command line

- authorize-security-group-ingress (AWS CLI)
- Grant-EC2SecurityGroupIngress (AWS Tools for Windows PowerShell)
To add one or more egress rules to a security group using the command line

- `authorize-security-group-egress` (AWS CLI)
- `Grant-EC2SecurityGroupEgress` (AWS Tools for Windows PowerShell)

Updating Security Group Rules

When you modify the protocol, port range, or source or destination of an existing security group rule using the console, the console deletes the existing rule and adds a new one for you.

To update a security group rule using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Security Groups.
3. Select the security group to update, and choose Inbound Rules to update a rule for inbound traffic or Outbound Rules to update a rule for outbound traffic.
4. Choose Edit. Modify the rule entry as required and choose Save.

To update the protocol, port range, or source or destination of an existing rule using the Amazon EC2 API or a command line tool, you cannot modify the rule. Instead, you must delete the existing rule and add a new rule. To update the rule description only, you can use the `update-security-group-rule-descriptions-ingress` and `update-security-group-rule-descriptions-egress` commands.

To update the description for an ingress security group rule using the command line

- `update-security-group-rule-descriptions-ingress` (AWS CLI)
- `Update-EC2SecurityGroupRuleIngressDescription` (AWS Tools for Windows PowerShell)

To update the description for an egress security group rule using the command line

- `update-security-group-rule-descriptions-egress` (AWS CLI)
- `Update-EC2SecurityGroupRuleEgressDescription` (AWS Tools for Windows PowerShell)

Deleting Rules from a Security Group

When you delete a rule from a security group, the change is automatically applied to any instances associated with the security group.

To delete a security group rule using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Security Groups.
3. Select a security group.
4. On the Inbound tab (for inbound rules) or Outbound tab (for outbound rules), choose Edit. Choose Delete (a cross icon) next to each rule to delete.
5. Choose Save.

To remove one or more ingress rules from a security group using the command line

- `revoke-security-group-ingress` (AWS CLI)
- `Revoke-EC2SecurityGroupIngress` (AWS Tools for Windows PowerShell)
To remove one or more egress rules from a security group using the command line

- `revoke-security-group-egress` (AWS CLI)
- `Revoke-EC2SecurityGroupEgress` (AWS Tools for Windows PowerShell)

Deleting a Security Group

You can't delete a security group that is associated with an instance. You can't delete the default security group. You can't delete a security group that is referenced by a rule in another security group in the same VPC. If your security group is referenced by one of its own rules, you must delete the rule before you can delete the security group.

To delete a security group using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Security Groups**.
3. Select a security group and choose **Actions, Delete Security Group**.
4. Choose **Yes, Delete**.

To delete a security group using the command line

- `delete-security-group` (AWS CLI)
- `Remove-EC2SecurityGroup` (AWS Tools for Windows PowerShell)

Security Group Rules Reference

You can create a security group and add rules that reflect the role of the instance that's associated with the security group. For example, an instance that's configured as a web server needs security group rules that allow inbound HTTP and HTTPS access, and a database instance needs rules that allow access for the type of database, such as access over port 3306 for MySQL.

The following are examples of the kinds of rules that you can add to security groups for specific kinds of access.

**Examples**

- Web Server Rules (p. 591)
- Database Server Rules (p. 592)
- Rules to Connect to Instances from Your Computer (p. 593)
- Rules to Connect to Instances from an Instance with the Same Security Group (p. 594)
- Rules for Path MTU Discovery (p. 594)
- Rules for Ping/ICMP (p. 594)
- DNS Server Rules (p. 595)
- Amazon EFS Rules (p. 595)
- Elastic Load Balancing Rules (p. 596)
- VPC Peering Rules (p. 597)

**Web Server Rules**

The following inbound rules allow HTTP and HTTPS access from any IP address. If your VPC is enabled for IPv6, you can add rules to control inbound HTTP and HTTPS traffic from IPv6 addresses.
### Protocol type | Protocol number | Port | Source IP | Notes |
---|---|---|---|---|
TCP | 6 | 80 (HTTP) | 0.0.0.0/0 | Allows inbound HTTP access from any IPv4 address |
TCP | 6 | 443 (HTTPS) | 0.0.0.0/0 | Allows inbound HTTPS access from any IPv4 address |
TCP | 6 | 80 (HTTP) | ::/0 | Allows inbound HTTP access from any IPv6 address |
TCP | 6 | 443 (HTTPS) | ::/0 | Allows inbound HTTPS access from any IPv6 address |

### Database Server Rules

The following inbound rules are examples of rules you might add for database access, depending on what type of database you're running on your instance. For more information about Amazon RDS instances, see the Amazon RDS User Guide.

For the source IP, specify one of the following:

- A specific IP address or range of IP addresses in your local network
- A security group ID for a group of instances that access the database

| Protocol type | Protocol number | Port | Notes |
---|---|---|---|
TCP | 6 | 1433 (MS SQL) | The default port to access a Microsoft SQL Server database, for example, on an Amazon RDS instance |
TCP | 6 | 3306 (MYSQL/Aurora) | The default port to access a MySQL or Aurora database, for example, on an Amazon RDS instance |
TCP | 6 | 5439 (Redshift) | The default port to access an Amazon Redshift cluster database. |
TCP | 6 | 5432 (PostgreSQL) | The default port to access a PostgreSQL database, for example, on an Amazon RDS instance |
### Protocol type | Protocol number | Port | Notes |
---|---|---|---|
TCP | 6 | 1521 (Oracle) | The default port to access an Oracle database, for example, on an Amazon RDS instance |

You can optionally restrict outbound traffic from your database servers, for example, if you want allow access to the Internet for software updates, but restrict all other kinds of traffic. You must first remove the default outbound rule that allows all outbound traffic.

### Protocol type | Protocol number | Port | Destination IP | Notes |
---|---|---|---|---|
TCP | 6 | 80 (HTTP) | 0.0.0.0/0 | Allows outbound HTTP access to any IPv4 address |
TCP | 6 | 443 (HTTPS) | 0.0.0.0/0 | Allows outbound HTTPS access to any IPv4 address |
TCP | 6 | 80 (HTTP) | ::/0 | (IPv6-enabled VPC only) Allows outbound HTTP access to any IPv6 address |
TCP | 6 | 443 (HTTPS) | ::/0 | (IPv6-enabled VPC only) Allows outbound HTTPS access to any IPv6 address |

### Rules to Connect to Instances from Your Computer

To connect to your instance, your security group must have inbound rules that allow SSH access (for Linux instances) or RDP access (for Windows instances).

### Protocol type | Protocol number | Port | Source IP |
---|---|---|---|
TCP | 6 | 22 (SSH) | The public IPv4 address of your computer, or a range of IP addresses in your local network. If your VPC is enabled for IPv6 and your instance has an IPv6 address, you can enter an IPv6 address or range. |
TCP | 6 | 3389 (RDP) | The public IPv4 address of your computer, or a range of IP addresses |
Rules to Connect to Instances from an Instance with the Same Security Group

To allow instances that are associated with the same security group to communicate with each other, you must explicitly add rules for this.

The following table describes the inbound rule for a security group that enables associated instances to communicate with each other. The rule allows all types of traffic.

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>Ports</th>
<th>Source IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1 (All)</td>
<td>-1 (All)</td>
<td>-1 (All)</td>
<td>The ID of the security group</td>
</tr>
</tbody>
</table>

Rules for Path MTU Discovery

The path MTU is the maximum packet size that’s supported on the path between the originating host and the receiving host. If a host sends a packet that's larger than the MTU of the receiving host or that's larger than the MTU of a device along the path, the receiving host returns the following ICMP message:

Destination Unreachable: Fragmentation Needed and Don't Fragment was Set

To ensure that your instance can receive this message and the packet does not get dropped, you must add an ICMP rule to your inbound security group rules.

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>ICMP type</th>
<th>ICMP code</th>
<th>Source IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICMP</td>
<td>1</td>
<td>3 (Destination Unreachable)</td>
<td>4 (Fragmentation Needed and Don't Fragment was Set)</td>
<td>The IP addresses of the hosts that communicate with your instance</td>
</tr>
</tbody>
</table>

Rules for Ping/ICMP

The ping command is a type of ICMP traffic. To ping your instance, you must add the following inbound ICMP rule.

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>ICMP type</th>
<th>ICMP code</th>
<th>Source IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICMP</td>
<td>1</td>
<td>8 (Echo)</td>
<td>N/A</td>
<td>The public IPv4 address of your instance</td>
</tr>
</tbody>
</table>
To use the `ping6` command to ping the IPv6 address for your instance, you must add the following inbound ICMPv6 rule.

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>ICMP type</th>
<th>ICMP code</th>
<th>Source IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICMPv6</td>
<td>58</td>
<td>128 (Echo)</td>
<td>0</td>
<td>The IPv6 address of your computer, or a range of IPv6 addresses in your local network</td>
</tr>
</tbody>
</table>

**DNS Server Rules**

If you've set up your EC2 instance as a DNS server, you must ensure that TCP and UDP traffic can reach your DNS server over port 53.

For the source IP, specify one of the following:

- An IP address or range of IP addresses in a network
- The ID of a security group for the set of instances in your network that require access to the DNS server

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>6</td>
<td>53</td>
</tr>
<tr>
<td>UDP</td>
<td>17</td>
<td>53</td>
</tr>
</tbody>
</table>

**Amazon EFS Rules**

If you're using an Amazon EFS file system with your Amazon EC2 instances, the security group that you associate with your Amazon EFS mount targets must allow traffic over the NFS protocol.

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>Ports</th>
<th>Source IP</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>6</td>
<td>2049 (NFS)</td>
<td>The ID of the security group.</td>
<td>Allows inbound NFS access from resources (including the mount target) associated with this security group.</td>
</tr>
</tbody>
</table>
To mount an Amazon EFS file system on your Amazon EC2 instance, you must connect to your instance. Therefore, the security group associated with your instance must have rules that allow inbound SSH from your local computer or local network.

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>Ports</th>
<th>Source IP</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>6</td>
<td>22 (SSH)</td>
<td>The IP address range of your local computer, or the range of IP addresses for your network.</td>
<td>Allows inbound SSH access from your local computer.</td>
</tr>
</tbody>
</table>

**Elastic Load Balancing Rules**

If you're using a load balancer, the security group associated with your load balancer must have rules that allow communication with your instances or targets.

<table>
<thead>
<tr>
<th>Protocol type</th>
<th>Protocol number</th>
<th>Port</th>
<th>Source IP</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>6</td>
<td>The listener port</td>
<td>For an Internet-facing load-balancer: 0.0.0.0/0 (all IPv4 addresses)</td>
<td>Allow inbound traffic on the load balancer listener port.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>For an internal load-balancer: the IPv4 CIDR block of the VPC</td>
<td></td>
</tr>
<tr>
<td>TCP</td>
<td>6</td>
<td>The instance listener port</td>
<td>The ID of the instance security group</td>
<td>Allow outbound traffic to instances on the instance listener port.</td>
</tr>
<tr>
<td>TCP</td>
<td>6</td>
<td>The health check port</td>
<td>The ID of the instance security group</td>
<td>Allow outbound traffic to instances on the health check port.</td>
</tr>
</tbody>
</table>

The security group rules for your instances must allow the load balancer to communicate with your instances on both the listener port and the health check port.
Controlling Access to Amazon EC2 Resources

Your security credentials identify you to services in AWS and grant you unlimited use of your AWS resources, such as your Amazon EC2 resources. You can use features of Amazon EC2 and AWS Identity and Access Management (IAM) to allow other users, services, and applications to use your Amazon EC2 resources without sharing your security credentials. You can use IAM to control how other users use resources in your AWS account, and you can use security groups to control access to your Amazon EC2 instances. You can choose to allow full use or limited use of your Amazon EC2 resources.

Contents

- Network Access to Your Instance (p. 597)
- Amazon EC2 Permission Attributes (p. 597)
- IAM and Amazon EC2 (p. 598)
- IAM Policies for Amazon EC2 (p. 599)
- IAM Roles for Amazon EC2 (p. 646)
- Authorizing Inbound Traffic for Your Windows Instances (p. 654)

Network Access to Your Instance

A security group acts as a firewall that controls the traffic allowed to reach one or more instances. When you launch an instance, you assign it one or more security groups. You add rules to each security group that control traffic for the instance. You can modify the rules for a security group at any time; the new rules are automatically applied to all instances to which the security group is assigned.

For more information, see Authorizing Inbound Traffic for Your Windows Instances (p. 654).

Amazon EC2 Permission Attributes

Your organization might have multiple AWS accounts. Amazon EC2 enables you to specify additional AWS accounts that can use your Amazon Machine Images (AMIs) and Amazon EBS snapshots. These

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Port</th>
<th>Action</th>
<th>Security Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>6</td>
<td>Allow traffic from the load balancer on the instance listener port.</td>
<td>The ID of the load balancer security group</td>
</tr>
<tr>
<td>TCP</td>
<td>6</td>
<td>Allow traffic from the load balancer on the health check port.</td>
<td>The ID of the load balancer security group</td>
</tr>
</tbody>
</table>

permissions work at the AWS account level only; you can't restrict permissions for specific users within
the specified AWS account. All users in the AWS account that you've specified can use the AMI or
snapshot.

Each AMI has a LaunchPermission attribute that controls which AWS accounts can access the AMI. For
more information, see Making an AMI Public (p. 55).

Each Amazon EBS snapshot has a createVolumePermission attribute that controls which AWS
accounts can use the snapshot. For more information, see Sharing an Amazon EBS Snapshot (p. 842).

IAM and Amazon EC2

IAM enables you to do the following:

- Create users and groups under your AWS account
- Assign unique security credentials to each user under your AWS account
- Control each user's permissions to perform tasks using AWS resources
- Allow the users in another AWS account to share your AWS resources
- Create roles for your AWS account and define the users or services that can assume them
- Use existing identities for your enterprise to grant permissions to perform tasks using AWS resources

By using IAM with Amazon EC2, you can control whether users in your organization can perform a task
using specific Amazon EC2 API actions and whether they can use specific AWS resources.

This topic helps you answer the following questions:

- How do I create groups and users in IAM?
- How do I create a policy?
- What IAM policies do I need to carry out tasks in Amazon EC2?
- How do I grant permissions to perform actions in Amazon EC2?
- How do I grant permissions to perform actions on specific resources in Amazon EC2?

Creating an IAM Group and Users

To create an IAM group

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Groups and then choose Create New Group.
3. For Group Name, type a name for your group, and then choose Next Step.
4. On the Attach Policy page, select an AWS managed policy and then choose Next Step. For example,
   for Amazon EC2, one of the following AWS managed policies might meet your needs:

   - PowerUserAccess
   - ReadOnlyAccess
   - AmazonEC2FullAccess
   - AmazonEC2ReadOnlyAccess
5. Choose Create Group.

Your new group is listed under Group Name.
To create an IAM user, add the user to your group, and create a password for the user

1. In the navigation pane, choose **Users, Add user**.
2. For **User name**, type a user name.
3. For **Access type**, select both **Programmatic access** and **AWS Management Console access**.
4. For **Console password**, choose one of the following:
   - **Autogenerated password**. Each user gets a randomly generated password that meets the current password policy in effect (if any). You can view or download the passwords when you get to the **Final page**.
   - **Custom password**. Each user is assigned the password that you type in the box.
5. Choose **Next: Permissions**.
6. On the **Set permissions** page, choose **Add user to group**. Select the check box next to the group that you created earlier and choose **Next: Review**.
7. Choose **Create user**.
8. To view the users' access keys (access key IDs and secret access keys), choose **Show** next to each password and secret access key to see. To save the access keys, choose **Download .csv** and then save the file to a safe location.
   
   **Important**
   You cannot retrieve the secret access key after you complete this step; if you misplace it you must create a new one.
9. Choose **Close**.
10. Give each user his or her credentials (access keys and password); this enables them to use services based on the permissions you specified for the IAM group.

Related Topics

For more information about IAM, see the following:

- IAM Policies for Amazon EC2 (p. 599)
- IAM Roles for Amazon EC2 (p. 646)
- AWS Identity and Access Management (IAM)
- IAM User Guide

IAM Policies for Amazon EC2

By default, IAM users don't have permission to create or modify Amazon EC2 resources, or perform tasks using the Amazon EC2 API. (This means that they also can't do so using the Amazon EC2 console or CLI.) To allow IAM users to create or modify resources and perform tasks, you must create IAM policies that grant IAM users permission to use the specific resources and API actions they'll need, and then attach those policies to the IAM users or groups that require those permissions.

When you attach a policy to a user or group of users, it allows or denies the users permission to perform the specified tasks on the specified resources. For more general information about IAM policies, see Permissions and Policies in the IAM User Guide. For more information about managing and creating custom IAM policies, see Managing IAM Policies.

Getting Started

An IAM policy must grant or deny permissions to use one or more Amazon EC2 actions. It must also specify the resources that can be used with the action, which can be all resources, or in some cases, specific resources. The policy can also include conditions that you apply to the resource.
Amazon EC2 partially supports resource-level permissions. This means that for some EC2 API actions, you cannot specify which resource a user is allowed to work with for that action; instead, you have to allow users to work with all resources for that action.

<table>
<thead>
<tr>
<th>Task</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understand the basic structure of a policy</td>
<td>Policy Syntax (p. 600)</td>
</tr>
<tr>
<td>Define actions in your policy</td>
<td>Actions for Amazon EC2 (p. 601)</td>
</tr>
<tr>
<td>Define specific resources in your policy</td>
<td>Amazon Resource Names for Amazon EC2 (p. 601)</td>
</tr>
<tr>
<td>Apply conditions to the use of the resources</td>
<td>Condition Keys for Amazon EC2 (p. 604)</td>
</tr>
<tr>
<td>Work with the available resource-level permissions for Amazon EC2</td>
<td>Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608)</td>
</tr>
<tr>
<td>Test your policy</td>
<td>Checking That Users Have the Required Permissions (p. 608)</td>
</tr>
<tr>
<td>Example policies for a CLI or SDK</td>
<td>Example Policies for Working with the AWS CLI or an AWS SDK (p. 611)</td>
</tr>
<tr>
<td>Example policies for the Amazon EC2 console</td>
<td>Example Policies for Working in the Amazon EC2 Console (p. 638)</td>
</tr>
</tbody>
</table>

**Policy Structure**

The following topics explain the structure of an IAM policy.

**Contents**
- Policy Syntax (p. 600)
- Actions for Amazon EC2 (p. 601)
- Amazon Resource Names for Amazon EC2 (p. 601)
- Condition Keys for Amazon EC2 (p. 604)
- Checking That Users Have the Required Permissions (p. 608)

**Policy Syntax**

An IAM policy is a JSON document that consists of one or more statements. Each statement is structured as follows:

```json
{
    "Statement": [{
        "Effect":"effect",
        "Action":"action",
        "Resource":"arn",
        "Condition":{
            "condition":{
                "key":"value"
            }
        }
    }
}
```
There are various elements that make up a statement:

- **Effect**: The effect can be Allow or Deny. By default, IAM users don't have permission to use resources and API actions, so all requests are denied. An explicit allow overrides the default. An explicit deny overrides any allows.

- **Action**: The action is the specific API action for which you are granting or denying permission. To learn about specifying action, see [Actions for Amazon EC2](#).

- **Resource**: The resource that's affected by the action. Some Amazon EC2 API actions allow you to include specific resources in your policy that can be created or modified by the action. To specify a resource in the statement, you need to use its Amazon Resource Name (ARN). For more information about specifying the ARN value, see [Amazon Resource Names for Amazon EC2](#). For more information about which API actions support which ARNs, see [Supported Resource-Level Permissions for Amazon EC2 API Actions](#). If the API action does not support ARNs, use the * wildcard to specify that all resources can be affected by the action.

- **Condition**: Conditions are optional. They can be used to control when your policy is in effect. For more information about specifying conditions for Amazon EC2, see [Condition Keys for Amazon EC2](#).

For more information about example IAM policy statements for Amazon EC2, see [Example Policies for Working with the AWS CLI or an AWS SDK](#).

### Actions for Amazon EC2

In an IAM policy statement, you can specify any API action from any service that supports IAM. For Amazon EC2, use the following prefix with the name of the API action: `ec2:`. For example: `ec2:RunInstances` and `ec2:CreateImage`.

To specify multiple actions in a single statement, separate them with commas as follows:

```
"Action": ["ec2:action1", "ec2:action2"]
```

You can also specify multiple actions using wildcards. For example, you can specify all actions whose name begins with the word "Describe" as follows:

```
"Action": "ec2:Describe*"
```

To specify all Amazon EC2 API actions, use the * wildcard as follows:

```
"Action": "ec2:*
```

For a list of Amazon EC2 actions, see [Actions in the Amazon EC2 API Reference](#).

### Amazon Resource Names for Amazon EC2

Each IAM policy statement applies to the resources that you specify using their ARNs.

**Important**

Currently, not all API actions support individual ARNs. We'll add support for additional API actions and ARNs for additional Amazon EC2 resources later. For information about which ARNs you can use with which Amazon EC2 API actions, as well as supported condition keys for each ARN, see [Supported Resource-Level Permissions for Amazon EC2 API Actions](#).

An ARN has the following general syntax:

```
arn:aws:[service]:[region]:[account]:resourceType/resourcePath
```
IAM Policies

**service**

The service (for example, `ec2`).

**region**

The region for the resource (for example, `us-east-1`).

**account**

The AWS account ID, with no hyphens (for example, `123456789012`).

**resourceType**

The type of resource (for example, `instance`).

**resourcePath**

A path that identifies the resource. You can use the * wildcard in your paths.

For example, you can indicate a specific instance (`i-1234567890abcdef0`) in your statement using its ARN as follows:

```
"Resource": "arn:aws:ec2:us-east-1:123456789012:instance/i-1234567890abcdef0"
```

You can also specify all instances that belong to a specific account by using the * wildcard as follows:

```
"Resource": "arn:aws:ec2:us-east-1:123456789012:instance/*"
```

To specify all resources, or if a specific API action does not support ARNs, use the * wildcard in the **Resource** element as follows:

```
"Resource": "*"
```

The following table describes the ARNs for each type of resource used by the Amazon EC2 API actions.

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>ARN</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Amazon EC2 resources</td>
<td><code>arn:aws:ec2:*</code></td>
</tr>
<tr>
<td>All Amazon EC2 resources owned by the specified account in the specified region</td>
<td><code>arn:aws:ec2:region:account:*</code></td>
</tr>
<tr>
<td></td>
<td>Where <code>cgw-id</code> is <code>cgw-xxxxxxxx</code></td>
</tr>
<tr>
<td></td>
<td>Where <code>dhcp-options-id</code> is <code>dopt-xxxxxxxx</code></td>
</tr>
<tr>
<td>Image</td>
<td><code>arn:aws:ec2:region::image/image-id</code></td>
</tr>
<tr>
<td></td>
<td>Where <code>image-id</code> is the ID of the AMI, AKI, or ARI, and <code>account</code> isn’t used</td>
</tr>
<tr>
<td>Instance</td>
<td><code>arn:aws:ec2:region:account:instance/instance-id</code></td>
</tr>
<tr>
<td>Resource Type</td>
<td>ARN</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Instance profile</td>
<td>arn:aws:iam::account:instance-profile/instance-profile-name</td>
</tr>
<tr>
<td></td>
<td>Where instance-profile-name is the name of the instance profile, and region isn’t used</td>
</tr>
<tr>
<td>Internet gateway</td>
<td>arn:aws:ec2:region:account:internet-gateway/igw-id</td>
</tr>
<tr>
<td></td>
<td>Where igw-id is igw-xxxxxxx</td>
</tr>
<tr>
<td>Key pair</td>
<td>arn:aws:ec2:region:account:key-pair/key-pair-name</td>
</tr>
<tr>
<td></td>
<td>Where key-pair-name is the key pair name (for example, gsg-keypair)</td>
</tr>
<tr>
<td></td>
<td>Where launch-template-id is lt-xxxxxxxxxxxxxxxxxxxx</td>
</tr>
<tr>
<td>NAT gateway</td>
<td>arn:aws:ec2:region:account:natgateway/natgateway-id</td>
</tr>
<tr>
<td></td>
<td>Where natgateway-id is nat-xxxxxxxxxxxxxxxxxxxx</td>
</tr>
<tr>
<td></td>
<td>Where nacl-id is acl-xxxxxxx</td>
</tr>
<tr>
<td></td>
<td>Where eni-id is eni-xxxxxxx</td>
</tr>
<tr>
<td>Placement group</td>
<td>arn:aws:ec2:region:account:placement-group/placement-group-name</td>
</tr>
<tr>
<td></td>
<td>Where placement-group-name is the placement group name (for example, my-cluster)</td>
</tr>
<tr>
<td>Reserved Instance</td>
<td>arn:aws:ec2:region:account:reserved-instances/reservation-id</td>
</tr>
<tr>
<td></td>
<td>Where reservation-id is xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx</td>
</tr>
<tr>
<td></td>
<td>Where route-table-id is rtb-xxxxxxx</td>
</tr>
<tr>
<td></td>
<td>Where security-group-id is sg-xxxxxxx</td>
</tr>
<tr>
<td>Snapshot</td>
<td>arn:aws:ec2:region::snapshot/snapshot-id</td>
</tr>
<tr>
<td></td>
<td>Where snapshot-id is snap-xxxxxxx or snap-xxxxxxxxxxxxxxxxxxxxx, and account isn’t used</td>
</tr>
<tr>
<td></td>
<td>Where spot-instance-request-id is sir-xxxxxxx</td>
</tr>
</tbody>
</table>
Many Amazon EC2 API actions involve multiple resources. For example, AttachVolume attaches an Amazon EBS volume to an instance, so an IAM user must have permissions to use the volume and the instance. To specify multiple resources in a single statement, separate their ARNs with commas, as follows:

```
"Resource": ["arn1", "arn2"]
```

For more general information about ARNs, see Amazon Resource Names (ARN) and AWS Service Namespaces in the Amazon Web Services General Reference. For more information about the resources that are created or modified by the Amazon EC2 actions, and the ARNs that you can use in your IAM policy statements, see Granting IAM Users Required Permissions for Amazon EC2 Resources in the Amazon EC2 API Reference.

### Condition Keys for Amazon EC2

In a policy statement, you can optionally specify conditions that control when it is in effect. Each condition contains one or more key-value pairs. Condition keys are not case-sensitive. We've defined AWS-wide condition keys, plus additional service-specific condition keys.

If you specify multiple conditions, or multiple keys in a single condition, we evaluate them using a logical AND operation. If you specify a single condition with multiple values for one key, we evaluate the condition using a logical OR operation. For permissions to be granted, all conditions must be met.

You can also use placeholders when you specify conditions. For example, you can grant an IAM user permission to use resources with a tag that specifies his or her IAM user name. For more information, see Policy Variables in the IAM User Guide.

**Important**

Many condition keys are specific to a resource, and some API actions use multiple resources. If you write a policy with a condition key, use the Resource element of the statement to specify the resource to which the condition key applies. If not, the policy may prevent users from performing the action at all, because the condition check fails for the resources to which
the condition key does not apply. If you do not want to specify a resource, or if you've written the Action element of your policy to include multiple API actions, then you must use the ...IfExists condition type to ensure that the condition key is ignored for resources that do not use it. For more information, see ...IfExists Conditions in the IAM User Guide.

Amazon EC2 implements the following service-specific condition keys. For information about which condition keys you can use with which Amazon EC2 resources, on an action-by-action basis, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608).

<table>
<thead>
<tr>
<th>Condition Key</th>
<th>Key-Value Pair</th>
<th>Evaluation Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>ec2:AccepterVpc</td>
<td>&quot;ec2:AccepterVpc&quot;:&quot;vpc-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where vpc-arn is the VPC ARN for the accepter VPC in a VPC peering connection</td>
<td></td>
</tr>
<tr>
<td>ec2:AuthorizedService</td>
<td>&quot;ec2:AuthorizedService&quot;:&quot;service-principal&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where service-principal is the service principal (for example, ecs.amazonaws.com)</td>
<td></td>
</tr>
<tr>
<td>ec2:AuthorizedUser</td>
<td>&quot;ec2:AuthorizedUser&quot;:&quot;principal-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where principal-arn is the ARN for the principal (for example, arn:aws:iam::123456789012:root)</td>
<td></td>
</tr>
<tr>
<td>ec2:AvailabilityZone</td>
<td>&quot;ec2:AvailabilityZone&quot;:&quot;az-api-name&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where az-api-name is the name of the Availability Zone (for example, us-east-2a)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>To list your Availability Zones, use describe-availability-zones</td>
<td></td>
</tr>
<tr>
<td>ec2:CreateAction</td>
<td>&quot;ec2:CreateAction&quot;:&quot;api-name&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where api-name is the name of the resource-creating action (for example, RunInstances)</td>
<td></td>
</tr>
<tr>
<td>ec2:EbsOptimized</td>
<td>&quot;ec2:EbsOptimized&quot;:&quot;optimized-flag&quot;</td>
<td>Boolean, Null</td>
</tr>
<tr>
<td></td>
<td>Where optimized-flag is true</td>
<td>false (for an instance)</td>
</tr>
<tr>
<td>ec2:ElasticGpuType</td>
<td>&quot;ec2:ElasticGpuType&quot;:&quot;elastic-gpu-type&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where elastic-gpu-type is the name of the elastic GPU type</td>
<td></td>
</tr>
<tr>
<td>ec2:Encrypted</td>
<td>&quot;ec2:Encrypted&quot;:&quot;encrypted-flag&quot;</td>
<td>Boolean, Null</td>
</tr>
<tr>
<td></td>
<td>Where encrypted-flag is true</td>
<td>false (for an EBS volume)</td>
</tr>
<tr>
<td>ec2:ImageType</td>
<td>&quot;ec2:ImageType&quot;:&quot;image-type-api-name&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where image-type-api-name is machine</td>
<td>aki</td>
</tr>
<tr>
<td>ec2:InstanceMarketType</td>
<td>&quot;ec2:InstanceMarketType&quot;:&quot;market-type&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where market-type is spot</td>
<td>on-demand</td>
</tr>
<tr>
<td>ec2:InstanceProfile</td>
<td>&quot;ec2:InstanceProfile&quot;:&quot;instance-profile-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where instance-profile-arn is the instance profile ARN</td>
<td></td>
</tr>
<tr>
<td>Condition Key</td>
<td>Key-Value Pair</td>
<td>Evaluation Types</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>ec2:InstanceType</td>
<td>&quot;instance-type-api-name&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where instance-type-api-name is the name of the instance type.</td>
<td></td>
</tr>
<tr>
<td>ec2:IsLaunchTemplate</td>
<td>&quot;launch-template-resource-flag&quot;</td>
<td>Boolean, Null</td>
</tr>
<tr>
<td></td>
<td>Where launch-template-resource-flag is true</td>
<td>false</td>
</tr>
<tr>
<td>ec2:LaunchTemplate</td>
<td>&quot;launch-template-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where launch-template-arn is the launch template ARN</td>
<td></td>
</tr>
<tr>
<td>ec2:Owner</td>
<td>&quot;account-id&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where account-id is amazon</td>
<td>aws-marketplace</td>
</tr>
<tr>
<td>ec2:ParentSnapshot</td>
<td>&quot;snapshot-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where snapshot-arn is the snapshot ARN</td>
<td></td>
</tr>
<tr>
<td>ec2:ParentVolume</td>
<td>&quot;volume-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where volume-arn is the volume ARN</td>
<td></td>
</tr>
<tr>
<td>ec2:Permission</td>
<td>&quot;permission&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where permission is INSTANCE-ATTACH</td>
<td>EIP-ASSOCIATE</td>
</tr>
<tr>
<td>ec2:PlacementGroup</td>
<td>&quot;placement-group-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where placement-group-arn is the placement group ARN</td>
<td></td>
</tr>
<tr>
<td>ec2:PlacementGroupStr</td>
<td>&quot;placement-group-strategy&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where placement-group-strategy is cluster</td>
<td>spread</td>
</tr>
<tr>
<td>ec2:ProductCode</td>
<td>&quot;product-code&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where product-code is the product code</td>
<td></td>
</tr>
<tr>
<td>ec2:Public</td>
<td>&quot;public-flag&quot;</td>
<td>Boolean, Null</td>
</tr>
<tr>
<td></td>
<td>Where public-flag is true</td>
<td>false (for an AMI)</td>
</tr>
<tr>
<td>ec2:Region</td>
<td>&quot;region-name&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where region-name is the name of the region (for example, us-east-2). To list your regions, use describe-regions. This condition key can be used with all Amazon EC2 actions.</td>
<td></td>
</tr>
<tr>
<td>ec2:RequesterVpc</td>
<td>&quot;vpc-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where vpc-arn is the VPC ARN for the requester VPC in a VPC peering connection</td>
<td></td>
</tr>
<tr>
<td>Condition Key</td>
<td>Key-Value Pair</td>
<td>Evaluation Types</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>ec2:ReservedInstancesOfferingType</td>
<td>&quot;offering-type&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where offering-type is No Upfront</td>
<td>Partial Upfront</td>
</tr>
<tr>
<td>ec2:ResourceTag/tag-key</td>
<td>&quot;tag-value&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where tag-key and tag-value are the tag-key pair</td>
<td></td>
</tr>
<tr>
<td>ec2:RootDeviceType/root-device-type-name</td>
<td></td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where root-device-type-name is ebs</td>
<td>instance-store</td>
</tr>
<tr>
<td>ec2:SnapshotTime/time</td>
<td></td>
<td>Date, Null</td>
</tr>
<tr>
<td></td>
<td>Where time is the snapshot creation time (for example, 2013-06-01T00:00:00Z)</td>
<td></td>
</tr>
<tr>
<td>ec2:Subnet</td>
<td>&quot;subnet-arn&quot;</td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where subnet-arn is the subnet ARN</td>
<td></td>
</tr>
<tr>
<td>ec2:Tenancy</td>
<td>&quot;tenancy-attribute&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where tenancy-attribute is default</td>
<td>dedicated</td>
</tr>
<tr>
<td>ec2:VolumeIops/volume-iops</td>
<td></td>
<td>Numeric, Null</td>
</tr>
<tr>
<td></td>
<td>Where volume-iops is the input/output operations per second (IOPS)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>For more information, see Amazon EBS Volume Types (p. 786).</td>
<td></td>
</tr>
<tr>
<td>ec2:VolumeSize/volume-size</td>
<td></td>
<td>Numeric, Null</td>
</tr>
<tr>
<td></td>
<td>Where volume-size is the size of the volume, in GiB</td>
<td></td>
</tr>
<tr>
<td>ec2:VolumeType/volume-type-name</td>
<td></td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where volume-type-name is gp2 for General Purpose SSD volumes, io1 for Provisioned IOPS SSD volumes, st1 for Throughput Optimized HDD volumes, sc1 for Cold HDD volumes, or standard for Magnetic volumes.</td>
<td></td>
</tr>
<tr>
<td>ec2:Vpc/vpc-arn</td>
<td></td>
<td>ARN, Null</td>
</tr>
<tr>
<td></td>
<td>Where vpc-arn is the VPC ARN</td>
<td></td>
</tr>
</tbody>
</table>

Amazon EC2 also implements the AWS-wide condition keys. For more information, see Information Available in All Requests in the IAM User Guide.

All Amazon EC2 actions support the aws:RequestedRegion and ec2:Region condition keys. For more information, see Example: Restricting Access to a Specific Region (p. 612).

The ec2:SourceInstanceARN key can be used for conditions that specify the ARN of the instance from which a request is made. This condition key is available AWS-wide and is not service-specific. For policy examples, see Allows an EC2 Instance to Attach or Detach Volumes and Example: Allowing a Specific Instance to View Resources in Other AWS Services (p. 637). The ec2:SourceInstanceARN key cannot be used as a variable to populate the ARN for the Resource element in a statement.
The following AWS condition keys were introduced for Amazon EC2 and are supported by a limited number of additional services.

<table>
<thead>
<tr>
<th>Condition Key</th>
<th>Key/Value Pair</th>
<th>Evaluation Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>aws:RequestTag/tag-key</td>
<td>&quot;aws:Request/tag-key&quot;: &quot;tag-value&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where tag-key and tag-value are the tag key-value pair</td>
<td></td>
</tr>
<tr>
<td>aws:TagKeys</td>
<td>&quot;aws:TagKeys&quot;: &quot;tag-key&quot;</td>
<td>String, Null</td>
</tr>
<tr>
<td></td>
<td>Where tag-key is a list of tag keys (for example, [&quot;A&quot;,&quot;B&quot;])</td>
<td></td>
</tr>
</tbody>
</table>

For example policy statements for Amazon EC2, see Example Policies for Working with the AWS CLI or an AWS SDK (p. 611).

Checking That Users Have the Required Permissions

After you’ve created an IAM policy, we recommend that you check whether it grants users the permissions to use the particular API actions and resources they need before you put the policy into production.

First, create an IAM user for testing purposes, and then attach the IAM policy that you created to the test user. Then, make a request as the test user.

If the Amazon EC2 action that you are testing creates or modifies a resource, you should make the request using the DryRun parameter (or run the AWS CLI command with the --dry-run option). In this case, the call completes the authorization check, but does not complete the operation. For example, you can check whether the user can terminate a particular instance without actually terminating it. If the test user has the required permissions, the request returns DryRunOperation; otherwise, it returns UnauthorizedOperation.

If the policy doesn’t grant the user the permissions that you expected, or is overly permissive, you can adjust the policy as needed and retest until you get the desired results.

**Important**

It can take several minutes for policy changes to propagate before they take effect. Therefore, we recommend that you allow five minutes to pass before you test your policy updates.

If an authorization check fails, the request returns an encoded message with diagnostic information. You can decode the message using the DecodeAuthorizationMessage action. For more information, see DecodeAuthorizationMessage in the AWS Security Token Service API Reference, and decode-authorization-message in the AWS CLI Command Reference.

Supported Resource-Level Permissions for Amazon EC2 API Actions

*Resource-level permissions* refers to the ability to specify which resources users are allowed to perform actions on. Amazon EC2 has partial support for resource-level permissions. This means that for certain Amazon EC2 actions, you can control when users are allowed to use those actions based on conditions that have to be fulfilled, or specific resources that users are allowed to use. For example, you can grant users permissions to launch instances, but only of a specific type, and only using a specific AMI.
If an Amazon EC2 API action does not support resource-level permissions, you can grant users permissions to use the action, but you have to specify a * for the resource element of your policy statement.

For more information about the resources that are created or modified by the Amazon EC2 actions, and the ARNs and Amazon EC2 condition keys that you can use in an IAM policy statement, see Actions, Resources, and Condition Keys for Amazon EC2 in the IAM User Guide.

For more information and for example policies, see IAM Policies for Amazon EC2 in the Amazon EC2 User Guide.

Resource-Level Permissions for Tagging

Some resource-creating Amazon EC2 API actions enable you to specify tags when you create the resource. For more information, see Tagging Your Resources (p. 966).

To enable users to tag resources on creation, they must have permissions to use the action that creates the resource (for example, ec2:RunInstances or ec2:CreateVolume). If tags are specified in the resource-creating action, Amazon performs additional authorization on the ec2:CreateTags action to verify if users have permissions to create tags. Therefore, users must also have explicit permissions to use the ec2:CreateTags action.

For the ec2:CreateTags action, you can use the ec2:CreateAction condition key to restrict tagging permissions to the resource-creating actions only. For example, the following policy allows users to launch instances and apply any tags to instances and volumes during launch. Users are not permitted to tag any existing resources (they cannot call the ec2:CreateTags action directly).

```json
{
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:RunInstances"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:region:account:*/*",
      "Condition": {
        "StringEquals": {
          "ec2:CreateAction" : "RunInstances"
        }
      }
    }
  ]
}
```

Similarly, the following policy allows users to create volumes and apply any tags to the volumes during volume creation. Users are not permitted to tag any existing resources (they cannot call the ec2:CreateTags action directly).

```json
{
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateVolume"
      ],
      "Resource": "*"
    }
  ]
}
```
IAM Policies

The `ec2:CreateTags` action is only evaluated if tags are applied during the resource-creating action. Therefore, a user that has permissions to create a resource (assuming there are no tagging conditions) does not require permissions to use the `ec2:CreateTags` action if no tags are specified in the request. However, if the user attempts to create a resource with tags, the request fails if the user does not have permissions to use the `ec2:CreateTags` action.

The `ec2:CreateTags` action is also evaluated if tags are provided in a launch template and the launch template is specified in the `ec2:RunInstances` action. For an example policy, see Tags in a Launch Template (p. 630).

You can control the tag keys and values that are applied to resources by using the following condition keys:

- `aws:RequestTag`: To indicate that a particular tag key or tag key and value must be present in a request. Other tags can also be specified in the request.
  - Use with the `StringEquals` condition operator to enforce a specific tag key and value combination, for example, to enforce the tag `cost-center` = `cc123`:
    
    ```
    "StringEquals": { "aws:RequestTag/cost-center": "cc123" }
    ```

- Use with the `StringLike` condition operator to enforce a specific tag key in the request; for example, to enforce the tag key `purpose`:
    
    ```
    "StringLike": { "aws:RequestTag/purpose": "*" }
    ```

- `aws:TagKeys`: To enforce the tag keys that are used in the request.
  - Use with the `ForAllValues` modifier to enforce specific tag keys if they are provided in the request (if tags are specified in the request, only specific tag keys are allowed; no other tags are allowed). For example, the tag keys `environment` or `cost-center` are allowed:
    
    ```
    "ForAllValues:StringEquals": { "aws:TagKeys": ["environment","cost-center"] }
    ```

  - Use with the `ForAnyValue` modifier to enforce the presence of at least one of the specified tag keys in the request. For example, at least one of the tag keys `environment` or `webserver` must be present in the request:
    
    ```
    "ForAnyValue:StringEquals": { "aws:TagKeys": ["environment","webserver"] }
    ```

These condition keys can be applied to resource-creating actions that support tagging, as well as the `ec2:CreateTags` and `ec2:DeleteTags` actions.
To force users to specify tags when they create a resource, you must use the `aws:RequestTag` condition key or the `aws:TagKeys` condition key with the `ForAnyValue` modifier on the resource-creating action. The `ec2:CreateTags` action is not evaluated if a user does not specify tags for the resource-creating action.

For conditions, the condition key is not case-sensitive and the condition value is case-sensitive. Therefore, to enforce the case-sensitivity of a tag key, use the `aws:TagKeys` condition key, where the tag key is specified as a value in the condition.

For more information about multi-value conditions, see Creating a Condition That Tests Multiple Key Values in the IAM User Guide. For example IAM policies, see Example Policies for Working with the AWS CLI or an AWS SDK (p. 611).

**Example Policies for Working with the AWS CLI or an AWS SDK**

The following examples show policy statements that you could use to control the permissions that IAM users have to Amazon EC2. These policies are designed for requests that are made with the AWS CLI or an AWS SDK. For example policies for working in the Amazon EC2 console, see Example Policies for Working in the Amazon EC2 Console (p. 638). For examples of IAM policies specific to Amazon VPC, see Controlling Access to Amazon VPC Resources.

**Examples**

- Example: Read-Only Access (p. 611)
- Example: Restricting Access to a Specific Region (p. 612)
- Working with Instances (p. 612)
- Working with Volumes (p. 614)
- Working with Snapshots (p. 616)
- Launching Instances (RunInstances) (p. 623)
- Example: Working with Reserved Instances (p. 633)
- Example: Tagging Resources (p. 634)
- Example: Working with IAM Roles (p. 636)
- Example: Working with Route Tables (p. 637)
- Example: Allowing a Specific Instance to View Resources in Other AWS Services (p. 637)
- Example: Working with Launch Templates (p. 638)

**Example: Read-Only Access**

The following policy grants users permissions to use all Amazon EC2 API actions whose names begin with `Describe`. The `Resource` element uses a wildcard to indicate that users can specify all resources with these API actions. The * wildcard is also necessary in cases where the API action does not support resource-level permissions. For more information about which ARNs you can use with which Amazon EC2 API actions, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608).

Users don't have permission to perform any actions on the resources (unless another statement grants them permission to do so) because they're denied permission to use API actions by default.

```json
{
    "Version": "2012-10-17",
    "Statement": [ {
        "Effect": "Allow",
        "Action": "ec2:Describe***",
        "Resource": "*"
    } ]
}
```
Example: Restricting Access to a Specific Region

The following policy denies users permission to use all Amazon EC2 API actions unless the Region is EU (Frankfurt). It uses the global condition key `aws:RequestedRegion`, which is supported by all Amazon EC2 API actions.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": "ec2:*",
      "Resource": "*",
      "Condition": {
        "StringNotEquals": {
          "aws:RequestedRegion": "eu-central-1"
        }
      }
    }
  ]
}
```

Alternatively, you can use the condition key `ec2:Region`, which is specific to Amazon EC2 and is supported by all Amazon EC2 API actions.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": "ec2:*",
      "Resource": "*",
      "Condition": {
        "StringNotEquals": {
          "ec2:Region": "eu-central-1"
        }
      }
    }
  ]
}
```

Working with Instances

**Examples**

- Example: Describe, Launch, Stop, Start, and Terminate All Instances (p. 612)
- Example: Describe All Instances, and Stop, Start, and Terminate Only Particular Instances (p. 613)

**Example: Describe, Launch, Stop, Start, and Terminate All Instances**

The following policy grants users permissions to use the API actions specified in the `Action` element. The `Resource` element uses a * wildcard to indicate that users can specify all resources with these API actions. The * wildcard is also necessary in cases where the API action does not support resource-level permissions. For more information about which ARNs you can use with which Amazon EC2 API actions, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608).
The users don't have permission to use any other API actions (unless another statement grants them permission to do so) because users are denied permission to use API actions by default.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeInstances", "ec2:DescribeImages",
        "ec2:DescribeKeyPairs", "ec2:DescribeSecurityGroups",
        "ec2:DescribeAvailabilityZones",
        "ec2:RunInstances", "ec2:TerminateInstances",
        "ec2:StopInstances", "ec2:StartInstances"
      ],
      "Resource": "*"
    }
  ]
}
```

**Example: Describe All Instances, and Stop, Start, and Terminate Only Particular Instances**

The following policy allows users to describe all instances, to start and stop only instances i-1234567890abcdef0 and i-0598c7d356eba48d7, and to terminate only instances in the US East (N. Virginia) Region (us-east-1) with the resource tag "purpose=test".

The first statement uses a * wildcard for the Resource element to indicate that users can specify all resources with the action; in this case, they can list all instances. The * wildcard is also necessary in cases where the API action does not support resource-level permissions (in this case, ec2:DescribeInstances). For more information about which ARNs you can use with which Amazon EC2 API actions, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608).

The second statement uses resource-level permissions for the StopInstances and StartInstances actions. The specific instances are indicated by their ARNs in the Resource element.

The third statement allows users to terminate all instances in the US East (N. Virginia) Region (us-east-1) that belong to the specified AWS account, but only where the instance has the tag "purpose=test". The Condition element qualifies when the policy statement is in effect.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": ["ec2:DescribeInstances"],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": ["ec2:StopInstances", "ec2:StartInstances"],
      "Resource": [
        "arn:aws:ec2:us-east-1:123456789012:instance/i-1234567890abcdef0",
        "arn:aws:ec2:us-east-1:123456789012:instance/i-0598c7d356eba48d7"
      ]
    },
    {
      "Effect": "Allow",
      "Action": ["ec2:TerminateInstances"],
      "Resource": "arn:aws:ec2:us-east-1:123456789012:instance/*",
```
Working with Volumes

Examples

- Example: Attaching and Detaching Volumes (p. 614)
- Example: Creating a Volume (p. 615)
- Example: Creating a Volume with Tags (p. 615)

Example: Attaching and Detaching Volumes

When an API action requires a caller to specify multiple resources, you must create a policy statement that allows users to access all required resources. If you need to use a Condition element with one or more of these resources, you must create multiple statements as shown in this example.

The following policy allows users to attach volumes with the tag "volume_user=iam-user-name" to instances with the tag "department=dev", and to detach those volumes from those instances. If you attach this policy to an IAM group, the aws:username policy variable gives each IAM user in the group permission to attach or detach volumes from the instances with a tag named volume_user that has his or her IAM user name as a value.

```json
{
"Version": "2012-10-17",
"Statement": [{
"Effect": "Allow",
"Action": [
"ec2:AttachVolume",
"ec2:DetachVolume"
],
"Resource": "arn:aws:ec2:us-east-1:123456789012:instance/**",
"Condition": {
"StringEquals": {
"ec2:ResourceTag/department": "dev"
}
}
},
{
"Effect": "Allow",
"Action": [
"ec2:AttachVolume",
"ec2:DetachVolume"
],
"Condition": {
"StringEquals": {
"ec2:ResourceTag/volume_user": "${aws:username}"
}
}
}
]
}
```
Example: Creating a Volume

The following policy allows users to use the CreateVolume API action. The user is allowed to create a volume only if the volume is encrypted and only if the volume size is less than 20 GiB.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": ["ec2:CreateVolume"],
      "Condition": {
        "NumericLessThan": {
          "ec2:VolumeSize": "20"
        },
        "Bool": {
          "ec2:Encrypted": "true"
        }
      }
    }
  ]
}
```

Example: Creating a Volume with Tags

The following policy includes the aws:RequestTag condition key that requires users to tag any volumes they create with the tags costcenter=115 and stack=prod. The aws:TagKeys condition key uses the ForAllValues modifier to indicate that only the keys costcenter and stack are allowed in the request (no other tags can be specified). If users don't pass these specific tags, or if they don't specify tags at all, the request fails.

For resource-creating actions that apply tags, users must also have permissions to use the CreateTags action. The second statement uses the ec2:CreateAction condition key to allow users to create tags only in the context of CreateVolume. Users cannot tag existing volumes or any other resources. For more information, see Resource-Level Permissions for Tagging (p. 609).

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowCreateTaggedVolumes",
      "Effect": "Allow",
      "Action": ["ec2:CreateVolume"],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/costcenter": "115",
          "aws:RequestTag/stack": "prod"
        },
        "ForAllValues:StringEquals": {
          "aws:TagKeys": ["costcenter","stack"]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": ["ec2:CreateTags"]
    }
  ]
}
```
The following policy allows users to create a volume without having to specify tags. The CreateTags action is only evaluated if tags are specified in the CreateVolume request. If users do specify tags, the tag must be purpose=test. No other tags are allowed in the request.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:CreateVolume",
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/purpose": "test",
          "ec2:CreateAction" : "CreateVolume"
        },
        "ForAllValues:StringEquals": {
          "aws:TagKeys": "purpose"
        }
      }
    }
  ]
}
```

Working with Snapshots

This section contains example policies for both CreateSnapshot (point-in-time snapshot of an EBS volume) and CreateSnapshots (multi-volume snapshots).

**Examples**

- Example: Creating a Snapshot (p. 616)
- Example: Creating Snapshots (p. 617)
- Example: Creating a Snapshot with Tags (p. 617)
- Example: Creating Snapshots with Tags (p. 618)
- Example: Modifying Permission Settings for Snapshots (p. 623)

**Example: Creating a Snapshot**

The following policy allows customers to use the CreateSnapshot API action. The customer can create snapshots only if the volume is encrypted and only if the volume size is less than 20 GiB.

```
{
  "Condition": {
    "StringEquals": {
      "ec2:CreateAction" : "CreateVolume"
    }
  }
}
```
Example: Creating Snapshots

The following policy allows customers to use the CreateSnapshots API action. The customer can create snapshots only if all of the volumes on the instance are type GP2.

```
"Version":"2012-10-17",
"Statement":[
  {
    "Effect":"Allow",
    "Action":"ec2:CreateSnapshots",
    "Resource":
    "arn:aws:ec2:us-east-1::snapshot/*",
    "arn:aws:ec2:*:*:instance/*"
  },
  {
    "Effect":"Allow",
    "Action":"ec2:CreateSnapshots",
    "Resource":
    "arn:aws:ec2:us-east-1:123456789012:volume/*",
    "Condition":{
      "NumericLessThan":{
        "ec2:VolumeSize":"20"
      },
      "Bool":{
        "ec2:Encrypted":"true"
      }
    }
  }
]
```

Example: Creating a Snapshot with Tags

The following policy includes the aws:RequestTag condition key that requires the customer to apply the tags costcenter=115 and stack=prod to any new snapshot. The aws:TagKeys condition key uses the ForAllValues modifier to indicate that only the keys costcenter and stack can be specified in the request. The request fails if either of these conditions is not met.

For resource-creating actions that apply tags, customers must also have permissions to use the CreateTags action. The third statement uses the ec2:CreateAction condition key to allow
customers to create tags only in the context of CreateSnapshot. Customers cannot tag existing
volumes or any other resources. For more information, see Resource-Level Permissions for Tagging.

```
{
  "Version":"2012-10-17",
  "Statement":[
    {
      "Effect":"Allow",
      "Action":"ec2:CreateSnapshot",
      "Resource":"arn:aws:ec2:us-east-1:123456789012:volume/*"
    },
    {
      "Sid":"AllowCreateTaggedSnapshots",
      "Effect":"Allow",
      "Action":"ec2:CreateSnapshot",
      "Resource":"arn:aws:ec2:us-east-1::snapshot/*",
      "Condition":{
        "StringEquals":{
          "aws:RequestTag/costcenter":"115",
          "aws:RequestTag/stack":"prod"
        },
        "ForAllValues:StringEquals":{
          "aws:TagKeys":[
            "costcenter",
            "stack"
          ]
        }
      }
    },
    {
      "Effect":"Allow",
      "Action":"ec2:CreateTags",
      "Resource":"arn:aws:ec2:us-east-1::snapshot/*",
      "Condition":{
        "StringEquals":{
          "ec2:CreateAction":"CreateSnapshot"
        }
      }
    }
  ]
}
```

Example: Creating Snapshots with Tags

The following policy includes the aws:RequestTag condition key that requires the customer to apply
the tags costcenter=115 and stack=prod to any new snapshot. The aws:TagKeys condition
key uses the ForAllValues modifier to indicate that only the keys costcenter and stack can be
specified in the request. The request fails if either of these conditions is not met.

```
{
  "Version":"2012-10-17",
  "Statement":[
    {
      "Effect":"Allow",
      "Action":"ec2:CreateSnapshots",
      "Resource":[
        "arn:aws:ec2:us-east-1::snapshot/*",
        "arn:aws:ec2:*:*:instance/*",
        "arn:aws:ec2:*:*:volume/*"
      ]
    },
    {
      "Effect":"Allow",
      "Action":"ec2:CreateTags",
      "Resource":[
        "arn:aws:ec2:us-east-1::snapshot/*",
        "arn:aws:ec2:*:*:instance/*",
        "arn:aws:ec2:*:*:volume/*"
      ]
    }
  ]
}
```
The following policy allows customers to create a snapshot without having to specify tags. The `CreateTags` action is evaluated only if tags are specified in the `CreateSnapshot` or `CreateSnapshots` request. If a tag is specified, the tag must be `purpose=test`. No other tags are allowed in the request.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:CreateSnapshot",
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": "ec2:CreateTags",
      "Resource": "arn:aws:ec2:us-east-1::snapshot/*",
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/costcenter": "115",
          "aws:RequestTag/stack": "prod"
        },
        "ForAllValues:StringEquals": {
          "aws:TagKeys": [
            "costcenter",
            "stack"
          ]
        }
      }
    }
  ]
}
```
IAM Policies

The following policy allows snapshots to be created only if the source volume is tagged with 
User:username for the customer, and the snapshot itself is tagged with Environment:Dev and User:username. The customer can add additional tags to the snapshot.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "ec2:CreateSnapshot",
            "Condition": {
                "StringEquals": {
                    "ec2:ResourceTag/User": "${aws:username}"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateSnapshot",
            "Resource": "arn:aws:ec2:us-east-1::snapshot/*",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/Environment": "Dev",
                    "aws:RequestTag/User": "${aws:username}"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateTags",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/purpose": "test",
                    "ec2:CreateAction": "CreateSnapshots"
                }
            }
        }
    ]
}
```

The following policy for CreateSnapshots allows snapshots to be created only if the source volume is tagged with User:username for the customer, and the snapshot itself is tagged with Environment:Dev and User:username.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "ec2:CreateSnapshot",
            "Condition": {
                "StringEquals": {
                    "ec2:ResourceTag/User": "${aws:username}"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateSnapshot",
            "Resource": "arn:aws:ec2:us-east-1::snapshot/*",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/Environment": "Dev",
                    "aws:RequestTag/User": "${aws:username}"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateTags",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/purpose": "test",
                    "ec2:CreateAction": "CreateSnapshots"
                }
            }
        }
    ]
}
```
The following policy allows deletion of a snapshot only if the snapshot is tagged with User:username for the customer.

```
{  
  "Version":"2012-10-17",
  "Statement":[
    {  
      "Effect":"Allow",
      "Action":"ec2:DeleteSnapshot",
      "Resource":"arn:aws:ec2:us-east-1::snapshot/*",
      "Condition":{
        "StringEquals":{
          "ec2:ResourceTag/User":"${aws:username}"  
        }
      }
    }
  ]
}
```

The following policy allows a customer to create a snapshot but denies the action if the snapshot being created has a tag key value=stack.

```
{  
  "Version":"2012-10-17",
  "Statement":[
    {  
      "Effect":"Allow",
      "Action":"ec2:CreateSnapshots",
      "Resource":"arn:aws:ec2:us-east-1:*:instance/**",
      "Condition":{
        "StringEquals":{
          "ec2:ResourceTag/User":"${aws:username}"  
        }
      }
    },
    {  
      "Effect":"Allow",
      "Action":"ec2:CreateSnapshots",
      "Resource":"arn:aws:ec2:us-east-1:123456789012:volume/*",
      "Condition":{
        "StringEquals":{
          "ec2:ResourceTag/User":"${aws:username}"  
        }
      }
    },
    {  
      "Effect":"Allow",
      "Action":"ec2:CreateSnapshots",
      "Resource":"arn:aws:ec2:us-east-1::snapshot/*",
      "Condition":{
        "StringEquals":{
          "aws:RequestTag/Environment":"Dev",
          "aws:RequestTag/User":"${aws:username}"  
        }
      }
    },
    {  
      "Effect":"Allow",
      "Action":"ec2:CreateTags",
      "Resource":"arn:aws:ec2:us-east-1::snapshot/*"  
    }
  ]
}
```
The following policy allows a customer to create snapshots but denies the action if the snapshots being created have a tag key value=stack.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSnapshots",
        "ec2:CreateTags"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Deny",
      "Action": "ec2:CreateSnapshot",
      "Resource": "arn:aws:ec2:us-east-1::snapshot/*",
      "Condition": {
        "ForAnyValue:StringEquals": {
          "aws:TagKeys": "stack"
        }
      }
    }
  ]
}
```

The following policy allows you to combine multiple actions into a single policy. You can only create a snapshot (in the context of CreateSnapshots) when the snapshot is created in Region us-east-1. You can only create snapshots (in the context of CreateSnapshots) when the snapshots are being created in the Region us-east-1 and when the instance type is t2*.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSnapshots",
        "ec2:CreateSnapshot",
        "ec2:CreateTags"
      ],
      "Resource": [
        "arn:aws:ec2:*:instance/*",
        "arn:aws:ec2:*:snapshot/*",
        "arn:aws:ec2:*:image/*"
      ]
    },
    {
      "Effect": "Deny",
      "Action": "ec2:CreateSnapshot",
      "Resource": "arn:aws:ec2:us-east-1::snapshot/*",
      "Condition": {
        "ForAnyValue:StringEquals": {
          "aws:TagKeys": "stack"
        }
      }
    }
  ]
}
```
IAM Policies

Example: Modifying Permission Settings for Snapshots

The following policy allows modification of a snapshot only if the snapshot is tagged with
User: username, where username is the customer's AWS account user name. The request fails if this condition is not met.

```
{
  "Version":"2012-10-17",
  "Statement":[
    {
      "Effect":"Allow",
      "Action":"ec2: ModifySnapshotAttribute",
      "Resource":"arn:aws:ec2:us-east-1::snapshot/*",
      "Condition":{
        "StringEquals":{
          "ec2:ResourceTag/user-name":"${aws:username}"}
      }]
  ]
}
```

Launching Instances (RunInstances)

The RunInstances API action launches one or more instances. RunInstances requires an AMI and creates an instance; and users can specify a key pair and security group in the request. Launching into a VPC requires a subnet, and creates a network interface. Launching from an Amazon EBS-backed AMI creates a volume. Therefore, the user must have permissions to use these Amazon EC2 resources. You can create a policy statement that requires users to specify an optional parameter on RunInstances, or restricts users to particular values for a parameter.

For more information about the resource-level permissions that are required to launch an instance, see Actions, Resources, and Condition Keys for Amazon EC2 in the IAM User Guide.

By default, users don't have permissions to describe, start, stop, or terminate the resulting instances. One way to grant the users permission to manage the resulting instances is to create a specific tag for each instance, and then create a statement that enables them to manage instances with that tag. For more information, see Working with Instances (p. 612).

Resources

- AMIs (p. 624)
- Instance Types (p. 625)
- Subnets (p. 626)
AMIs

The following policy allows users to launch instances using only the specified AMIs, ami-9e1670f7 and ami-45cf5c3c. The users can't launch an instance using other AMIs (unless another statement grants the users permission to do so).

```json
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Action": "ec2:RunInstances",
        "Resource": [
            "arn:aws:ec2::image/ami-9e1670f7",
            "arn:aws:ec2::image/ami-45cf5c3c",
            "arn:aws:ec2::account:instance/**",
            "arn:aws:ec2::account:volume/**",
            "arn:aws:ec2::account:key-pair/**",
            "arn:aws:ec2::account:security-group/**",
            "arn:aws:ec2::account:subnet/**",
            "arn:aws:ec2::account:network-interface/**"
        ]
    },
    {
        "Effect": "Allow",
        "Action": "ec2:RunInstances",
        "Resource": [
            "arn:aws:ec2::account:instance/*",
            "arn:aws:ec2::account:subnet/*",
            "arn:aws:ec2::account:volume/*",
            "arn:aws:ec2::account:network-interface/*",
            "arn:aws:ec2::account:security-group/*"
        ]
    }
}
```

Alternatively, the following policy allows users to launch instances from all AMIs owned by Amazon. The `Condition` element of the first statement tests whether `ec2:Owner` is `amazon`. The users can't launch an instance using other AMIs (unless another statement grants the users permission to do so).

```json
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Action": "ec2:RunInstances",
        "Resource": [
            "arn:aws:ec2::image/ami-*"
        ],
        "Condition": {
            "StringEquals": {
                "ec2:Owner": "amazon"
            }
        }
    },
    {
        "Effect": "Allow",
        "Action": "ec2:RunInstances",
        "Resource": [
            "arn:aws:ec2::account:instance/**",
            "arn:aws:ec2::account:volume/**",
            "arn:aws:ec2::account:network-interface/**",
            "arn:aws:ec2::account:key-pair/**",
            "arn:aws:ec2::account:security-group/**"
        ]
    }
}
```
Instance Types

The following policy allows users to launch instances using only the \texttt{t2.micro} or \texttt{t2.small} instance type, which you might do to control costs. The users can't launch larger instances because the \texttt{Condition} element of the first statement tests whether \texttt{ec2:InstanceType} is either \texttt{t2.micro} or \texttt{t2.small}.


```
{
   "Version": "2012-10-17",
   "Statement": [{
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
         "arn:aws:ec2:region:account:instance/*"
      ],
      "Condition": {
         "StringEquals": {
            "ec2:InstanceType": ["t2.micro", "t2.small"]
         }
      }
   },
   "Effect": "Allow",
   "Action": "ec2:RunInstances",
   "Resource": [
      "arn:aws:ec2:region:image/ami-*",
      "arn:aws:ec2:region:account:subnet/**",
      "arn:aws:ec2:region:account:volume/**",
      "arn:aws:ec2:region:account:security-group/**",
      "arn:aws:ec2:region:account:security-group/**"
   ]
}
```

Alternatively, you can create a policy that denies users permissions to launch any instances except \texttt{t2.micro} and \texttt{t2.small} instance types.

```
{
   "Version": "2012-10-17",
   "Statement": [{
      "Effect": "Deny",
      "Action": "ec2:RunInstances",
      "Resource": [
         "arn:aws:ec2:region:account:instance/*"
      ],
      "Condition": {
         "StringNotEquals": {
            "ec2:InstanceType": ["t2.micro", "t2.small"]
         }
      }
   },
   "Effect": "Allow",
   "Action": "ec2:RunInstances",
   "Resource": [
      "arn:aws:ec2:region:image/ami-*",
      "arn:aws:ec2:region:account:subnet/**",
      "arn:aws:ec2:region:account:volume/**",
      "arn:aws:ec2:region:account:security-group/**",
      "arn:aws:ec2:region:account:security-group/**"
   ]
}
```
Subnets

The following policy allows users to launch instances using only the specified subnet, subnet-12345678. The group can’t launch instances into any another subnet (unless another statement grants the users permission to do so).

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Action": "ec2:RunInstances",
        "Resource": [
            "arn:aws:ec2:region:account:instance/*",
            "arn:aws:ec2:region:account:volume/*",
            "arn:aws:ec2:region:account:image/ami-*",
            "arn:aws:ec2:region:account:key-pair/*",
            "arn:aws:ec2:region:account:security-group/*"
        ]
    }]
}
```

Alternatively, you could create a policy that denies users permissions to launch an instance into any other subnet. The statement does this by denying permission to create a network interface, except where subnet subnet-12345678 is specified. This denial overrides any other policies that are created to allow launching instances into other subnets.

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Deny",
        "Action": "ec2:RunInstances",
        "Resource": [
        ],
        "Condition": {
            "ArnNotEquals": {
            }
        }
    },
    {
        "Effect": "Allow",
        "Action": "ec2:RunInstances",
        "Resource": [
            "arn:aws:ec2:region:image/ami-*",
            "arn:aws:ec2:region:account:instance/*",
            "arn:aws:ec2:region:account:subnet/*",
            "arn:aws:ec2:region:account:volume/*",
            "arn:aws:ec2:region:account:key-pair/*",
        ]
    }]
}
```
IAM Policies

EBS Volumes

The following policy allows users to launch instances only if the EBS volumes for the instance are encrypted. The user must launch an instance from an AMI that was created with encrypted snapshots, to ensure that the root volume is encrypted. Any additional volume that the user attaches to the instance during launch must also be encrypted.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
        "arn:aws:ec2:*:*:volume/*"
      ],
      "Condition": {
        "Bool": {
          "ec2:Encrypted": "true"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
        "arn:aws:ec2:*::*:image/ami-*",
        "arn:aws:ec2:*::*:network-interface/*",
        "arn:aws:ec2:*::*:instance/*",
        "arn:aws:ec2:*::*:subnet/*",
        "arn:aws:ec2:*::*:key-pair/*",
        "arn:aws:ec2:*::*:security-group/*"
      ]
    }
  ]
}
```

Tags

The following policy allows users to launch instances and tag the instances during creation. For resource-creating actions that apply tags, users must have permissions to use the `CreateTags` action. The second statement uses the `ec2:CreateAction` condition key to allow users to create tags only in the context of `RunInstances`, and only for instances. Users cannot tag existing resources, and users cannot tag volumes using the `RunInstances` request.

For more information, see Resource-Level Permissions for Tagging (p. 609).

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": "*
    }
  ]
}
```
The following policy includes the `aws:RequestTag` condition key that requires users to tag any instances and volumes that are created by `RunInstances` with the tags `environment=production` and `purpose=webserver`. The `aws:TagKeys` condition key uses the `ForAllValues` modifier to indicate that only the keys `environment` and `purpose` are allowed in the request (no other tags can be specified). If no tags are specified in the request, the request fails.
The following policy uses the `ForAnyValue` modifier on the `aws:TagKeys` condition to indicate that at least one tag must be specified in the request, and it must contain the key `environment` or `webserver`. The tag must be applied to both instances and volumes. Any tag values can be specified in the request.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [ "ec2:CreateAction" : "RunInstances"
            ],
            "Resource": [
                "arn:aws:ec2:region::image/*",
                "arn:aws:ec2:region:account:subnet/*",
                "arn:aws:ec2:region:account:security-group/*",
                "arn:aws:ec2:region:account:key-pair/*"
            ],
            "Condition": {
                "ForAnyValue:StringEquals": {
                    "aws:TagKeys": ["environment","webserver"]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [ "ec2:CreateAction" : "RunInstances"
            ],
            "Resource": [
                "arn:aws:ec2:region:account:volume/*",
                "arn:aws:ec2:region:account:instance/*"
            ],
            "Condition": {
                "StringEquals": {
                    "ec2:CreateAction" : "RunInstances"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [ "ec2:CreateTags"
            ],
            "Resource": "arn:aws:ec2:region:account:*/**",
            "Condition": {
                "StringEquals": {
                    "ec2:CreateAction" : "RunInstances"
                }
            }
        }
    ]
}
```
Tags in a Launch Template

In the following example, users can launch instances, but only if they use a specific launch template (lt-09477bcd97b0d310e). The ec2:IsLaunchTemplateResource condition key prevents users from overriding any of the resources specified in the launch template. The second part of the statement allows users to tag instances on creation—this part of the statement is necessary if tags are specified for the instance in the launch template.

```json
"Version": "2012-10-17",
"Statement": [
  {
    "Effect": "Allow",
    "Action": ["ec2:RunInstances"],
    "Resource": "*"
  },
  {
    "Effect": "Allow",
    "Action": ["ec2:CreateTags"],
    "Resource": "arn:aws:ec2:region:account:*/*",
    "Condition": {
      "StringEquals": {
        "aws:RequestTag/purpose": "test",
        "ec2:CreateAction": "RunInstances"
      },
      "ForAllValues:StringEquals": {
        "aws:TagKeys": "purpose"
      }
    }
  }
]
```
Elastic GPUs

In the following policy, users can launch an instance and specify an elastic GPU to attach to the instance. Users can launch instances in any region, but they can only attach an elastic GPU during a launch in the `us-east-2` region.

The `ec2:ElasticGpuType` condition key uses the `ForAnyValue` modifier to indicate that only the elastic GPU types `eg1.medium` and `eg1.large` are allowed in the request.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": ["ec2:RunInstances"],
      "Resource": [
        "arn:aws:ec2::*:account:elastic-gpu/**"
      ],
      "Condition": {
        "StringEquals": {
          "ec2:Region": "us-east-2"
        },
        "ForAnyValue:StringLike": {
          "ec2:ElasticGpuType": [
            "eg1.medium",
            "eg1.large"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
        "arn:aws:ec2::*:image/ami-*",
        "arn:aws:ec2::*:account:network-interface/**",
        "arn:aws:ec2::*:account:instance/**",
        "arn:aws:ec2::*:account:subnet/**",
        "arn:aws:ec2::*:account:volume/**",
        "arn:aws:ec2::*:account:key-pair/**",
        "arn:aws:ec2::*:account:security-group/**"
      ]
    }
  ]
}
```

Launch Templates

In the following example, users can launch instances, but only if they use a specific launch template (`lt-09477bcd97b0d310e`). Users can override any parameters in the launch template by specifying the parameters in the `RunInstances` action.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
        "arn:aws:ec2::*:image/ami-*",
        "arn:aws:ec2::*:account:network-interface/**",
        "arn:aws:ec2::*:account:instance/**",
        "arn:aws:ec2::*:account:subnet/**",
        "arn:aws:ec2::*:account:volume/**",
        "arn:aws:ec2::*:account:key-pair/**",
        "arn:aws:ec2::*:account:security-group/**"
      ]
    }
  ]
}
```
In this example, users can launch instances only if they use a launch template. The policy uses the `ec2:IsLaunchTemplateResource` condition key to prevent users from overriding any of the launch template resources in the `RunInstances` request.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Condition": {
        "ArnLike": {
          "ec2:LaunchTemplate": "arn:aws:ec2:region:account:launch-template/*"},
        "Bool": {
          "ec2:IsLaunchTemplateResource": "true"
        }
      }
    }
  ]
}
```

The following example policy allows user to launch instances, but only if they use a launch template. Users cannot override the subnet and network interface parameters in the request; these parameters can only be specified in the launch template. The first part of the statement uses the `NotResource` element to allow all other resources except subnets and network interfaces. The second part of the statement allows the subnet and network interface resources, but only if they are sourced from the launch template.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Condition": {
        "ArnLike": {
          "ec2:LaunchTemplate": "arn:aws:ec2:region:account:launch-template/*"}
      }
    },
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": ["arn:aws:ec2:region:account:subnet/*",
      "Condition": {
                    "ArnLike": {
                      "ec2:LaunchTemplate": "arn:aws:ec2:region:account:launch-template/*"}
                    }
    }
  ]
}
```
The following example allows users to launch instances only if they use a launch template, and only if the launch template has the tag `Purpose=Webservers`. Users cannot override any of the launch template parameters in the `RunInstances` action.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Condition": {
        "ArnLike": {
          "ec2:LaunchTemplate": "arn:aws:ec2:region:account:launch-template/**
        },
        "Bool": {
          "ec2:IsLaunchTemplateResource": "true"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Condition": {
        "StringEquals": {
          "ec2:ResourceTag/Purpose": "Webservers"
        }
      }
    }
  ]
}
```

**Example: Working with Reserved Instances**

The following policy gives users permission to view, modify, and purchase Reserved Instances in your account.

It is not possible to set resource-level permissions for individual Reserved Instances. This policy means that users have access to all the Reserved Instances in the account.

The `Resource` element uses a * wildcard to indicate that users can specify all resources with the action; in this case, they can list and modify all Reserved Instances in the account. They can also purchase Reserved instances using the account credentials. The * wildcard is also necessary in cases where the API action does not support resource-level permissions.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": 
```
To allow users to view and modify the Reserved Instances in your account, but not purchase new Reserved Instances.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeReservedInstances",
                "ec2:ModifyReservedInstances",
                "ec2:DescribeAvailabilityZones"
            ],
            "Resource": "*"
        }
    ]
}
```

Example: Tagging Resources

The following policy allows users to use the `CreateTags` action to apply tags to an instance only if the tag contains the key `environment` and the value `production`. The `ForAllValues` modifier is used with the `aws:TagKeys` condition key to indicate that only the key `environment` is allowed in the request (no other tags are allowed). The user cannot tag any other resource types.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateTags"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/environment": "production"
                },
                "ForAllValues:StringEquals": {
                    "aws:TagKeys": [
                        "environment"
                    ]
                }
            }
        }
    ]
}
```

The following policy allows users to tag any taggable resource that already has a tag with a key of `owner` and a value of the IAM username. In addition, users must specify a tag with a key of `environment` and a value of either `test` or `prod` in the request. Users can specify additional tags in the request.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:PutResourceTaggingPolicy"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/environment": "production"
                },
                "ForAllValues:StringEquals": {
                    "aws:TagKeys": [
                        "environment"
                    ]
                }
            }
        }
    ]
}
```
You can create an IAM policy that allows users to delete specific tags for a resource. For example, the following policy allows users to delete tags for a volume if the tag keys specified in the request are `environment` or `cost-center`. Any value can be specified for the tag but the tag key must match either of the specified keys.

**Note**
If you delete a resource, all tags associated with the resource are also deleted. Users do not need permissions to use the `ec2:DeleteTags` action to delete a resource that has tags; they only need permissions to perform the deleting action.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "ec2:DeleteTags",
            "Condition": {
                "ForAllValues:StringEquals": {
                    "aws:TagKeys": ["environment","cost-center"]
                }
            }
        }
    ]
}
```

This policy allows users to delete only the `environment=prod` tag on any resource, and only if the resource is already tagged with a key of `owner` and a value of the IAM username. Users cannot delete any other tags for a resource.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": ["ec2:DeleteTags"],
            "Resource": "arn:aws:ec2:region:account:*/*",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/environment": "prod",
                    "ec2:ResourceTag/owner": "\${aws:username}"}
            },
            "ForAllValues:StringEquals": {
```
Example: Working with IAM Roles

The following policy allows users to attach, replace, and detach an IAM role to instances that have the tag department=test. Replacing or detaching an IAM role requires an association ID, therefore the policy also grants users permission to use the ec2:DescribeIamInstanceProfileAssociations action.

IAM users must have permission to use the iam:PassRole action in order to pass the role to the instance.

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:AssociateIamInstanceProfile",
        "ec2:ReplaceIamInstanceProfileAssociation",
        "ec2:DisassociateIamInstanceProfile"
      ],
      "Condition": {
        "StringEquals": {
          "ec2:ResourceTag/department": "test"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": "ec2:DescribeIamInstanceProfileAssociations",
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": "iam:PassRole",
      "Resource": "*"
    }
  ]
}
```

The following policy allows users to attach or replace an IAM role for any instance. Users can only attach or replace IAM roles with names that begin with TestRole-. For the iam:PassRole action, ensure that you specify the name of the IAM role and not the instance profile (if the names are different). For more information, see Instance Profiles (p. 647).

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:AssociateIamInstanceProfile",
        "ec2:ReplaceIamInstanceProfileAssociation"
      ],
      "Resource": "*"
    }
  ]
}
```
Example: Working with Route Tables

The following policy allows users to add, remove, and replace routes for route tables that are associated with VPC vpc-ec43eb89 only. To specify a VPC for the ec2:Vpc condition key, you must specify the full ARN of the VPC.

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": [  
                "ec2:DeleteRoute",  
                "ec2:CreateRoute",  
                "ec2:ReplaceRoute"  
            ],  
            "Resource": [  
            ],  
            "Condition": {  
                "StringEquals": {  
                    "ec2:Vpc": "arn:aws:ec2:region:account:vpc/vpc-ec43eb89"  
                }  
            }  
        }  
    ]  
}
```

Example: Allowing a Specific Instance to View Resources in Other AWS Services

The following is an example of a policy that you might attach to an IAM role. The policy allows an instance to view resources in various AWS services. It uses the ec2:SourceInstanceARN condition key to specify that the instance from which the request is made must be instance i-093452212644b0dd6. If the same IAM role is associated with another instance, the other instance cannot perform any of these actions.

The ec2:SourceInstanceARN key is an AWS-wide condition key, therefore it can be used for other service actions, not just Amazon EC2.

```
{  
    "Version": "2012-10-17",  
    "Statement": [  
        {  
            "Effect": "Allow",  
            "Action": [  
                "ec2:DescribeVolumes",  
                "s3:ListAllMyBuckets",  
                "dynamodb:ListTables",  
            ]  
        }  
    ]  
}
```


Example: Working with Launch Templates

The following policy allows users to create a launch template version and modify a launch template, but only for a specific launch template (lt-09477bcd97b0d3abc). Users cannot work with other launch templates.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "ec2:CreateLaunchTemplateVersion",
                "ec2:ModifyLaunchTemplate"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:ec2:region:account:launch-template/lt-09477bcd97b0d3abc"
        }
    ]
}
```

The following policy allows users to delete any launch template and launch template version, provided that the launch template has the tag Purpose=Testing.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "ec2:DeleteLaunchTemplate",
                "ec2:DeleteLaunchTemplateVersions"
            ],
            "Effect": "Allow",
            "Condition": {
                "StringEquals": {
                    "ec2:ResourceTag/Purpose": "Testing"
                }
            }
        }
    ]
}
```

Example Policies for Working in the Amazon EC2 Console

You can use IAM policies to grant users permissions to view and work with specific resources in the Amazon EC2 console. You can use the example policies in the previous section; however, they are
designated for requests that are made with the AWS CLI or an AWS SDK. The console uses additional API actions for its features, so these policies may not work as expected. For example, a user that has permission to use only the DescribeVolumes API action will encounter errors when trying to view volumes in the console. This section demonstrates policies that enable users to work with specific parts of the console.

**Tip**
To help you work out which API actions are required to perform tasks in the console, you can use a service such as AWS CloudTrail. For more information, see the [AWS CloudTrail User Guide](https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/cloudtrail-aws-monitoring.html). If your policy does not grant permission to create or modify a specific resource, the console displays an encoded message with diagnostic information. You can decode the message using the DecodeAuthorizationMessage API action for AWS STS, or the `decode-authorization-message` command in the AWS CLI.

**Examples**
- Example: Read-Only Access (p. 639)
- Example: Using the EC2 Launch Wizard (p. 640)
- Example: Working with Volumes (p. 642)
- Example: Working with Security Groups (p. 643)
- Example: Working with Elastic IP Addresses (p. 645)
- Example: Working with Reserved Instances (p. 645)

For additional information about creating policies for the Amazon EC2 console, see the following AWS Security Blog post: [Granting Users Permission to Work in the Amazon EC2 Console](https://aws.amazon.com/blogs/security/)

**Example: Read-Only Access**

To allow users to view all resources in the Amazon EC2 console, you can use the same policy as the following example: Example: Read-Only Access (p. 611). Users cannot perform any actions on those resources or create new resources, unless another statement grants them permission to do so.

**View instances, AMIs, and snapshots**

Alternatively, you can provide read-only access to a subset of resources. To do this, replace the * wildcard in the `ec2:Describe*` API action with specific `ec2:Describe*` actions for each resource. The following policy allows users to view all instances, AMIs, and snapshots in the Amazon EC2 console. The `ec2:DescribeTags` action allows users to view public AMIs. The console requires the tagging information to display public AMIs; however, you can remove this action to allow users to view only private AMIs.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeInstances", "ec2:DescribeImages",
                "ec2:DescribeTags", "ec2:DescribeSnapshots"
            ],
            "Resource": "*
        }
    ]
}
```

**Note**
The Amazon EC2 `ec2:Describe*` API actions do not support resource-level permissions, so you cannot control which individual resources users can view in the console. Therefore, the *
wildcard is necessary in the `Resource` element of the above statement. For more information about which ARNs you can use with which Amazon EC2 API actions, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608).

View instances and CloudWatch metrics

The following policy allows users to view instances in the Amazon EC2 console, as well as CloudWatch alarms and metrics in the Monitoring tab of the Instances page. The Amazon EC2 console uses the CloudWatch API to display the alarms and metrics, so you must grant users permission to use the `cloudwatch:DescribeAlarms` and `cloudwatch:GetMetricStatistics` actions.

```json
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": [
            "ec2:DescribeInstances",
            "cloudwatch:DescribeAlarms",
            "cloudwatch:GetMetricStatistics"
         ],
         "Resource": "*"
      }
   ]
}
```

Example: Using the EC2 Launch Wizard

The Amazon EC2 launch wizard is a series of screens with options to configure and launch an instance. Your policy must include permission to use the API actions that allow users to work with the wizard's options. If your policy does not include permission to use those actions, some items in the wizard cannot load properly, and users cannot complete a launch.

Basic launch wizard access

To complete a launch successfully, users must be given permission to use the `ec2:RunInstances` API action, and at least the following API actions:

- `ec2:DescribeImages`: To view and select an AMI.
- `ec2:DescribeVpcs`: To view the available network options.
- `ec2:DescribeSubnets`: To view all available subnets for the chosen VPC.
- `ec2:DescribeSecurityGroups`: To view the security groups page in the wizard. Users can select an existing security group.
- `ec2:DescribeKeyPairs` or `ec2:CreateKeyPair`: To select an existing key pair, or create a new one.

```json
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": [
            "ec2:DescribeInstances", "ec2:DescribeImages",
            "ec2:DescribeKeyPairs","ec2:DescribeVpcs", "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups"
         ],
         "Resource": "*"
      },
      {
         "Effect": "Allow",
         "Action": [
            "ec2:RunInstances",
            "ec2:DescribeImages", "ec2:DescribeKeyPairs","ec2:DescribeVpcs", "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups"
         ],
         "Resource": "*"
      }
   ]
}
```
You can add API actions to your policy to provide more options for users, for example:

- `ec2:DescribeAvailabilityZones`: To view and select a specific Availability Zone.
- `ec2:DescribeNetworkInterfaces`: To view and select existing network interfaces for the selected subnet.
- `ec2:CreateSecurityGroup`: To create a new security group; for example, to create the wizard's suggested `launch-wizard-x` security group. However, this action alone only creates the security group; it does not add or modify any rules. To add inbound rules, users must be granted permission to use the `ec2:AuthorizeSecurityGroupIngress` API action. To add outbound rules to VPC security groups, users must be granted permission to use the `ec2:AuthorizeSecurityGroupEgress` API action. To modify or delete existing rules, users must be granted permission to use the relevant `ec2:RevokeSecurityGroup*` API action.
- `ec2:CreateTags`: To tag the resources that are created by `RunInstances`. For more information, see Resource-Level Permissions for Tagging (p. 609). If users do not have permission to use this action and they attempt to apply tags on the tagging page of the launch wizard, the launch fails.

**Important**
Be careful about granting users permission to use the `ec2:CreateTags` action. This limits your ability to use the `ec2:ResourceTag` condition key to restrict the use of other resources; users can change a resource's tag in order to bypass those restrictions.

Currently, the Amazon EC2 Describe* API actions do not support resource-level permissions, so you cannot restrict which individual resources users can view in the launch wizard. However, you can apply resource-level permissions on the `ec2:RunInstances` API action to restrict which resources users can use to launch an instance. The launch fails if users select options that they are not authorized to use.

**Restrict access to specific instance type, subnet, and region**

The following policy allows users to launch `m1.small` instances using AMIs owned by Amazon, and only into a specific subnet (`subnet-1a2b3c4d`). Users can only launch in the sa-east-1 region. If users select a different region, or select a different instance type, AMI, or subnet in the launch wizard, the launch fails.

The first statement grants users permission to view the options in the launch wizard, as demonstrated in the example above. The second statement grants users permission to use the network interface, volume, key pair, security group, and subnet resources for the `ec2:RunInstances` action, which are required to launch an instance into a VPC. For more information about using the `ec2:RunInstances` action, see Launching Instances (RunInstances) (p. 623). The third and fourth statements grant users permission to use the instance and AMI resources respectively, but only if the instance is an `m1.small` instance, and only if the AMI is owned by Amazon.
Example: Working with Volumes

The following policy grants users permission to view and create volumes, and attach and detach volumes to specific instances.

Users can attach any volume to instances that have the tag "purpose=test", and also detach volumes from those instances. To attach a volume using the Amazon EC2 console, it is helpful for users to have permission to use the `ec2:DescribeInstances` action, as this allows them to select an instance from a pre-populated list in the **Attach Volume** dialog box. However, this also allows users to view all instances on the **Instances** page in the console, so you can omit this action.

In the first statement, the `ec2:DescribeAvailabilityZones` action is necessary to ensure that a user can select an Availability Zone when creating a volume.

Users cannot tag the volumes that they create (either during or after volume creation).
Example: Working with Security Groups

**View security groups and add and remove rules**

The following policy grants users permission to view security groups in the Amazon EC2 console, and to add and remove inbound and outbound rules for existing security groups that have the tag `Department=Test`.

In the first statement, the `ec2:DescribeTags` action allows users to view tags in the console, which makes it easier for users to identify the security groups that they are allowed to modify.
Working with the Create Security Group dialog box

You can create a policy that allows users to work with the Create Security Group dialog box in the Amazon EC2 console. To use this dialog box, users must be granted permission to use at least the following API actions:

- `ec2:CreateSecurityGroup`: To create a new security group.
- `ec2:DescribeVpcs`: To view a list of existing VPCs in the VPC list.

With these permissions, users can create a new security group successfully, but they cannot add any rules to it. To work with rules in the Create Security Group dialog box, you can add the following API actions to your policy:

- `ec2:AuthorizeSecurityGroupIngress`: To add inbound rules.
- `ec2:AuthorizeSecurityGroupEgress`: To add outbound rules to VPC security groups.
- `ec2:RevokeSecurityGroupIngress`: To modify or delete existing inbound rules. This is useful to allow users to use the Copy to new feature in the console. This feature opens the Create Security Group dialog box and populates it with the same rules as the security group that was selected.
- `ec2:RevokeSecurityGroupEgress`: To modify or delete outbound rules for VPC security groups. This is useful to allow users to modify or delete the default outbound rule that allows all outbound traffic.
- `ec2:DeleteSecurityGroup`: To cater for when invalid rules cannot be saved. The console first creates the security group, and then adds the specified rules. If the rules are invalid, the action fails, and the console attempts to delete the security group. The user remains in the Create Security Group dialog box so that they can correct the invalid rule and try to create the security group again. This API action is not required, but if a user is not granted permission to use it and attempts to create a security group with invalid rules, the security group is created without any rules, and the user must add them afterward.

Currently, the `ec2:CreateSecurityGroup` API action does not support resource-level permissions; however, you can apply resource-level permissions to the `ec2:AuthorizeSecurityGroupIngress` and `ec2:AuthorizeSecurityGroupEgress` actions to control how users can create rules.

The following policy grants users permission to use the Create Security Group dialog box, and to create inbound and outbound rules for security groups that are associated with a specific VPC (`vpc-1a2b3c4d`). Users can create security groups for EC2-Classic or another VPC, but they cannot add any rules to them. Similarly, users cannot add any rules to any existing security group that’s not associated with VPC `vpc-1a2b3c4d`. Users are also granted permission to view all security groups in the console. This makes it easier for users to identify the security groups to which they can add inbound rules. This policy also grants users permission to delete security groups that are associated with VPC `vpc-1a2b3c4d`.

```json
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "ec2:DescribeSecurityGroups", "ec2:CreateSecurityGroup", "ec2:DescribeVpcs"
    ],
    "Resource": "*"
  },
  {
    "Effect": "Allow",
    "Action": [
      "ec2:DeleteSecurityGroup", "ec2:AuthorizeSecurityGroupIngress",
      "ec2:AuthorizeSecurityGroupEgress"
    ],
    "Resource": "*"
  }
}
```
Example: Working with Elastic IP Addresses

To allow users to view Elastic IP addresses in the Amazon EC2 console, you must grant users permission to use the `ec2:DescribeAddresses` action.

To allow users to work with Elastic IP addresses, you can add the following actions to your policy.

- `ec2:AllocateAddress`: To allocate an Elastic IP address.
- `ec2:ReleaseAddress`: To release an Elastic IP address.
- `ec2:AssociateAddress`: To associate an Elastic IP address with an instance or a network interface.
- `ec2:DescribeNetworkInterfaces` and `ec2:DescribeInstances`: To work with the **Associate address** screen. The screen displays the available instances or network interfaces to which you can associate an Elastic IP address.
- `ec2:DisassociateAddress`: To disassociate an Elastic IP address from an instance or a network interface.

The following policy allows users to view, allocate, and associate Elastic IP addresses with instances. Users cannot associate Elastic IP addresses with network interfaces, disassociate Elastic IP addresses, or release them.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeAddresses",
                "ec2:AllocateAddress",
                "ec2:DescribeInstances",
                "ec2:AssociateAddress"
            ],
            "Resource": "*
        }
    ]
}
```

Example: Working with Reserved Instances

The following policy can be attached to an IAM user. It gives the user access to view and modify Reserved Instances in your account, as well as purchase new Reserved Instances in the AWS Management Console.

This policy allows users to view all the Reserved Instances, as well as On-Demand Instances, in the account. It’s not possible to set resource-level permissions for individual Reserved Instances.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeInstances",
                "ec2:PurchaseReservedInstancesOffering",
                "ec2:AcceptReservedInstancesOffering",
                "ec2:AcceptPartialInstancesFromReservedInstancesOffering"
            ],
            "Resource": "*
        }
    ]
}
```
IAM Roles for Amazon EC2

Applications must sign their API requests with AWS credentials. Therefore, if you are an application developer, you need a strategy for managing credentials for your applications that run on EC2 instances. For example, you can securely distribute your AWS credentials to the instances, enabling the applications on those instances to use your credentials to sign requests, while protecting your credentials from other users. However, it's challenging to securely distribute credentials to each instance, especially those that AWS creates on your behalf, such as Spot Instances or instances in Auto Scaling groups. You must also be able to update the credentials on each instance when you rotate your AWS credentials.

We designed IAM roles so that your applications can securely make API requests from your instances, without requiring you to manage the security credentials that the applications use. Instead of creating and distributing your AWS credentials, you can delegate permission to make API requests using IAM roles as follows:

1. Create an IAM role.
2. Define which accounts or AWS services can assume the role.
3. Define which API actions and resources the application can use after assuming the role.
4. Specify the role when you launch your instance, or attach the role to an existing instance.
5. Have the application retrieve a set of temporary credentials and use them.

For example, you can use IAM roles to grant permissions to applications running on your instances that need to use a bucket in Amazon S3. You can specify permissions for IAM roles by creating a policy in JSON format. These are similar to the policies that you create for IAM users. If you change a role, the change is propagated to all instances.

You cannot attach multiple IAM roles to a single instance, but you can attach a single IAM role to multiple instances. For more information about creating and using IAM roles, see Roles in the IAM User Guide.

You can apply resource-level permissions to your IAM policies to control the users’ ability to attach, replace, or detach IAM roles for an instance. For more information, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608) and the following example: Example: Working with IAM Roles (p. 636).

Topics
- Instance Profiles (p. 647)
Instance Profiles

Amazon EC2 uses an instance profile as a container for an IAM role. When you create an IAM role using the IAM console, the console creates an instance profile automatically and gives it the same name as the role to which it corresponds. If you use the Amazon EC2 console to launch an instance with an IAM role or to attach an IAM role to an instance, you choose the role based on a list of instance profile names.

If you use the AWS CLI, API, or an AWS SDK to create a role, you create the role and instance profile as separate actions, with potentially different names. If you then use the AWS CLI, API, or an AWS SDK to launch an instance with an IAM role or to attach an IAM role to an instance, specify the instance profile name.

An instance profile can contain only one IAM role. This limit cannot be increased.

For more information, see Instance Profiles in the IAM User Guide.

Retrieving Security Credentials from Instance Metadata

An application on the instance retrieves the security credentials provided by the role from the instance metadata item `iam/security-credentials/role-name`. The application is granted the permissions for the actions and resources that you've defined for the role through the security credentials associated with the role. These security credentials are temporary and we rotate them automatically. We make new credentials available at least five minutes before the expiration of the old credentials.

**Warning**

If you use services that use instance metadata with IAM roles, ensure that you don't expose your credentials when the services make HTTP calls on your behalf. The types of services that could expose your credentials include HTTP proxies, HTML/CSS validator services, and XML processors that support XML inclusion.

The following command retrieves the security credentials for an IAM role named `s3access`.

```
```

The following is example output.

```json
{
"Code" : "Success",
"LastUpdated" : "2012-04-26T16:39:16Z",
"Type" : "AWS-HMAC",
"AccessKeyId" : "ASIAIOSFODNN7EXAMPLE",
"SecretAccessKey" : "wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY",
"Token" : "token",
"Expiration" : "2017-05-17T15:09:54Z"
}
```

For applications, AWS CLI, and Tools for Windows PowerShell commands that run on the instance, you do not have to explicitly get the temporary security credentials — the AWS SDKs, AWS CLI, and Tools for Windows PowerShell automatically get the credentials from the EC2 instance metadata service and use them. To make a call outside of the instance using temporary security credentials (for example, to test IAM policies), you must provide the access key, secret key, and the session token. For more information, see Using Temporary Security Credentials to Request Access to AWS Resources in the IAM User Guide.
Granting an IAM User Permission to Pass an IAM Role to an Instance

To enable an IAM user to launch an instance with an IAM role or to attach or replace an IAM role for an existing instance, you must grant the user permission to pass the role to the instance.

The following IAM policy grants users permission to launch instances (ec2:RunInstances) with an IAM role, or to attach or replace an IAM role for an existing instance (ec2:AssociateIamInstanceProfile and ec2:ReplaceIamInstanceProfileAssociation).

```json
{
   "Version": "2012-10-17",
   "Statement": [ 
      {
         "Effect": "Allow",
         "Action": [ 
            "ec2:RunInstances",
            "ec2:AssociateIamInstanceProfile",
            "ec2:ReplaceIamInstanceProfileAssociation"
         ],
         "Resource": "*"
      },
      {
         "Effect": "Allow",
         "Action": "iam:PassRole",
         "Resource": "*"
      }
   ]
}
```

This policy grants IAM users access to all your roles by specifying the resource as "*" in the policy. However, consider whether users who launch instances with your roles (ones that exist or that you create later on) might be granted permissions that they don't need or shouldn't have.

Working with IAM Roles

You can create an IAM role and attach it to an instance during or after launch. You can also replace or detach an IAM role for an instance.

Contents
- Creating an IAM Role (p. 648)
- Launching an Instance with an IAM Role (p. 650)
- Attaching an IAM Role to an Instance (p. 651)
- Replacing an IAM Role (p. 652)
- Detaching an IAM Role (p. 653)

Creating an IAM Role

You must create an IAM role before you can launch an instance with that role or attach it to an instance.

To create an IAM role using the IAM console

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Roles, Create role.
3. On the **Select role type** page, choose **EC2** and the **EC2** use case. Choose **Next: Permissions**.
4. On the **Attach permissions policy** page, select an AWS managed policy that grants your instances access to the resources that they need.
5. On the **Review** page, type a name for the role and choose **Create role**.

Alternatively, you can use the AWS CLI to create an IAM role.

**To create an IAM role and instance profile (AWS CLI)**

- Create an IAM role with a policy that allows the role to use an Amazon S3 bucket.
  
  a. Create the following trust policy and save it in a text file named `ec2-role-trust-policy.json`.

  ```json
  {
  "Version": "2012-10-17",
  "Statement": [
  {
  "Effect": "Allow",
  "Principal": { "Service": "ec2.amazonaws.com"},
  "Action": "sts:AssumeRole"
  }
  ]
  }
  ```

  b. Create the `s3access` role and specify the trust policy that you created.

  ```bash
  aws iam create-role --role-name s3access --assume-role-policy-document file://ec2-role-trust-policy.json
  ```

  ```json
  "RoleId": "AROAIIZKPBKS2LEXAMPLE",
  "CreateDate": "2013-12-12T23:46:37.247Z",
  "RoleName": "s3access",
  "Path": "/",
  "Arn": "arn:aws:iam::123456789012:role/s3access"
  }
  ```

  c. Create an access policy and save it in a text file named `ec2-role-access-policy.json`. For example, this policy grants administrative permissions for Amazon S3 to applications running on the instance.

  ```json
  {
  "Version": "2012-10-17",
  "Statement": [
  {
  "Effect": "Allow",
  "Action": ["s3:*"]
  }
  ]
  }
  ```
d. Attach the access policy to the role.

```bash
aws iam put-role-policy --role-name s3access --policy-name S3-Permissions --policy-document file://ec2-role-access-policy.json
```

e. Create an instance profile named `s3access-profile`.

```bash
aws iam create-instance-profile --instance-profile-name s3access-profile
{
  "InstanceProfile": {
    "InstanceProfileId": "AIPAJTLBPJLEGREXAMPLE",
    "Roles": [],
    "CreateDate": "2013-12-12T23:53:34.093Z",
    "InstanceProfileName": "s3access-profile",
    "Path": "/",
    "Arn": "arn:aws:iam::123456789012:instance-profile/s3access-profile"
  }
}
```

f. Add the `s3access` role to the `s3access-profile` instance profile.

```bash
aws iam add-role-to-instance-profile --instance-profile-name s3access-profile --role-name s3access
```

For more information about these commands, see `create-role`, `put-role-policy`, and `create-instance-profile` in the AWS CLI Command Reference.

Alternatively, you can use the following AWS Tools for Windows PowerShell commands:

- New-IAMRole
- Register-IAMRolePolicy
- New-IAMInstanceProfile

### Launching an Instance with an IAM Role

After you've created an IAM role, you can launch an instance, and associate that role with the instance during launch.

**Important**

After you create an IAM role, it may take several seconds for the permissions to propagate. If your first attempt to launch an instance with a role fails, wait a few seconds before trying again. For more information, see Troubleshooting Working with Roles in the IAM User Guide.

**To launch an instance with an IAM role (console)**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. On the dashboard, choose **Launch Instance**.
3. Select an AMI and instance type and then choose **Next: Configure Instance Details**.
4. On the **Configure Instance Details** page, for **IAM role**, select the IAM role that you created.

**Note**

The IAM role list displays the name of the instance profile that you created when you created your IAM role. If you created your IAM role using the console, the instance profile
was created for you and given the same name as the role. If you created your IAM role using the AWS CLI, API, or an AWS SDK, you may have named your instance profile differently.

5. Configure any other details, then follow the instructions through the rest of the wizard, or choose Review and Launch to accept default settings and go directly to the Review Instance Launch page.

6. Review your settings, then choose Launch to choose a key pair and launch your instance.

7. If you are using the Amazon EC2 API actions in your application, retrieve the AWS security credentials made available on the instance and use them to sign the requests. The AWS SDK does this for you.

```
```

Alternatively, you can use the AWS CLI to associate a role with an instance during launch. You must specify the instance profile in the command.

**To launch an instance with an IAM role (AWS CLI)**

1. Use the `run-instances` command to launch an instance using the instance profile. The following example shows how to launch an instance with the instance profile.

```
aws ec2 run-instances --image-id ami-11aa22bb --iam-instance-profile Name="s3access-profile" --key-name my-key-pair --security-groups my-security-group --subnet-id subnet-1a2b3c4d
```

Alternatively, use the New-EC2Instance Tools for Windows PowerShell command.

2. If you are using the Amazon EC2 API actions in your application, retrieve the AWS security credentials made available on the instance and use them to sign the requests. The AWS SDK does this for you.

```
```

**Attaching an IAM Role to an Instance**

To attach an IAM role to an instance that has no role, the instance can be in the stopped or running state.

**To attach an IAM role to an instance (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance, choose Actions, Instance Settings, Attach/Replace IAM role.
4. Select the IAM role to attach to your instance, and choose Apply.

**To attach an IAM role to an instance (AWS CLI)**

1. If required, describe your instances to get the ID of the instance to which to attach the role.

```
aws ec2 describe-instances
```

2. Use the `associate-iam-instance-profile` command to attach the IAM role to the instance by specifying the instance profile. You can use the Amazon Resource Name (ARN) of the instance profile, or you can use its name.
Replacing an IAM Role

To replace the IAM role on an instance that already has an attached IAM role, the instance must be in the running state. You can do this if you want to change the IAM role for an instance without detaching the existing one first; for example, to ensure that API actions performed by applications running on the instance are not interrupted.

To replace an IAM role for an instance (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance, choose Actions, Instance Settings, Attach/Replace IAM role.
4. Select the IAM role to attach to your instance, and choose Apply.

To replace an IAM role for an instance (AWS CLI)

1. If required, describe your IAM instance profile associations to get the association ID for the IAM instance profile to replace.

```bash
aws ec2 describe-iam-instance-profile-associations
```

2. Use the replace-iam-instance-profile-association command to replace the IAM instance profile by specifying the association ID for the existing instance profile and the ARN or name of the instance profile that should replace it.

```bash
aws ec2 replace-iam-instance-profile-association --association-id iip-assoc-0044d817db60d4ba --iam-instance-profile Name="TestRole-2"
```

Alternatively, use the following Tools for Windows PowerShell commands:

- Get-EC2Instance
- Register-EC2IamInstanceProfile
Alternatively, use the following Tools for Windows PowerShell commands:

- Get-EC2IamInstanceProfileAssociation
- Set-EC2IamInstanceProfileAssociation

**Detaching an IAM Role**

You can detach an IAM role from a running or stopped instance.

**To detach an IAM role from an instance (console)**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Select the instance, choose **Actions, Instance Settings, Attach/Replace IAM role**.
4. For **IAM role**, choose **No Role**. Choose **Apply**.
5. In the confirmation dialog box, choose **Yes, Detach**.

**To detach an IAM role from an instance (AWS CLI)**

1. If required, use describe-iam-instance-profile-associations to describe your IAM instance profile associations and get the association ID for the IAM instance profile to detach.

```bash
aws ec2 describe-iam-instance-profile-associations
{
   "IamInstanceProfileAssociations": [
      {
         "InstanceId": "i-088ce778fbeb4361",
         "State": "associated",
         "AssociationId": "iip-assoc-0044d817db6c0a4ba",
         "IamInstanceProfile": {
            "Id": "AIPAJEDNCAA64SSD265D6",
            "Arn": "arn:aws:iam::123456789012:instance-profile/TestRole-2"
         }
      }
   ]
}
```

2. Use the disassociate-iam-instance-profile command to detach the IAM instance profile using its association ID.

```bash
aws ec2 disassociate-iam-instance-profile --association-id iip-assoc-0044d817db6c0a4ba
{
   "IamInstanceProfileAssociation": {
      "InstanceId": "i-087711ddaf98f9489",
      "State": "disassociating",
      "AssociationId": "iip-assoc-0044d817db6c0a4ba",
      "IamInstanceProfile": {
         "Id": "AIPAJEDNCAA64SSD265D6",
         "Arn": "arn:aws:iam::123456789012:instance-profile/TestRole-2"
      }
   }
}
```
Alternatively, use the following Tools for Windows PowerShell commands:

- Get-EC2IamInstanceProfileAssociation
- Unregister-EC2IamInstanceProfile

**Authorizing Inbound Traffic for Your Windows Instances**

Security groups enable you to control traffic to your instance, including the kind of traffic that can reach your instance. For example, you can allow computers from only your home network to access your instance using RDP. If your instance is a web server, you can allow all IP addresses to access your instance using HTTP or HTTPS, so that external users can browse the content on your web server.

Your default security groups and newly created security groups include default rules that do not enable you to access your instance from the Internet. For more information, see Default Security Groups (p. 586) and Custom Security Groups (p. 587). To enable network access to your instance, you must allow inbound traffic to your instance. To open a port for inbound traffic, add a rule to a security group that you associated with your instance when you launched it.

To connect to your instance, you must set up a rule to authorize RDP traffic from your computer's public IPv4 address. To allow RDP traffic from additional IP address ranges, add another rule for each range you need to authorize.

If you've enabled your VPC for IPv6 and launched your instance with an IPv6 address, you can connect to your instance using its IPv6 address instead of a public IPv4 address. Your local computer must have an IPv6 address and must be configured to use IPv6.

If you need to enable network access to a Linux instance, see Authorizing Inbound Traffic for Your Linux Instances in the Amazon EC2 User Guide for Linux Instances.

**Before You Start**

Decide who requires access to your instance; for example, a single host or a specific network that you trust such as your local computer's public IPv4 address. The security group editor in the Amazon EC2 console can automatically detect the public IPv4 address of your local computer for you. Alternatively, you can use the search phrase "what is my IP address" in an internet browser, or use the following service: Check IP. If you are connecting through an ISP or from behind your firewall without a static IP address, you need to find out the range of IP addresses used by client computers.

**Warning**

If you use 0.0.0.0/0, you enable all IPv4 addresses to access your instance using RDP. If you use ::/0, you enable all IPv6 address to access your instance. This is acceptable for a short time in a test environment, but it's unsafe for production environments. In production, you authorize only a specific IP address or range of addresses to access your instance.

Windows Firewall may also block incoming traffic. If you're having trouble setting up access to your instance, you may have to disable Windows Firewall. For more information, see Remote Desktop Can't Connect to the Remote Computer (p. 1045).

**Adding a Rule for Inbound RDP Traffic to a Windows Instance**

Security groups act as a firewall for associated instances, controlling both inbound and outbound traffic at the instance level. You must add rules to a security group that enable you to connect to your Windows instance from your IP address using RDP.
To add a rule to a security group for inbound RDP traffic over IPv4 (console)

1. In the navigation pane of the Amazon EC2 console, choose **Instances**. Select your instance and look at the **Description** tab; **Security groups** lists the security groups that are associated with the instance. Choose **view inbound rules** to display a list of the rules that are in effect for the instance.

2. In the navigation pane, choose **Security Groups**. Select one of the security groups associated with your instance.

3. In the details pane, on the **Inbound** tab, choose **Edit**. In the dialog, choose **Add Rule**, and then choose **RDP** from the **Type** list.

4. In the **Source** field, choose **My IP** to automatically populate the field with the public IPv4 address of your local computer. Alternatively, choose **Custom** and specify the public IPv4 address of your computer or network in CIDR notation. For example, if your IPv4 address is 203.0.113.25, specify 203.0.113.25/32 to list this single IPv4 address in CIDR notation. If your company allocates addresses from a range, specify the entire range, such as 203.0.113.0/24.

For information about finding your IP address, see *Before You Start (p. 654).*

5. Choose **Save**.

If you launched an instance with an IPv6 address and want to connect to your instance using its IPv6 address, you must add rules that allow inbound IPv6 traffic over RDP.

To add a rule to a security group for inbound RDP traffic over IPv6 (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.

2. In the navigation pane, choose **Security Groups**. Select the security group for your instance.

3. Choose **Inbound, Edit, Add Rule**.

4. For **Type**, choose RDP.

5. In the **Source** field, specify the IPv6 address of your computer in CIDR notation. For example, if your IPv6 address is 2001:db8:1234:1a00:9691:9503:25ad:1761, specify 2001:db8:1234:1a00:9691:9503:25ad:1761/128 to list the single IPv6 address in CIDR notation. If your company allocates addresses from a range, specify the entire range, such as 2001:db8:1234:1a00::/64.

6. Choose **Save**.

**Note**

Be sure to run the following commands on your local system, not on the instance itself. For more information about these command line interfaces, see *Accessing Amazon EC2 (p. 3).*

To add a rule to a security group using the command line

1. Find the security group that is associated with your instance using one of the following commands:
   
   - **describe-instance-attribute** (AWS CLI)

   ```bash
   aws ec2 describe-instance-attribute --instance-id instance_id --attribute groupSet
   ```

   - **Get-EC2InstanceAttribute** (AWS Tools for Windows PowerShell)

   ```powershell
   PS C:\> (Get-EC2InstanceAttribute -InstanceId instance_id -Attribute groupSet).Groups
   ```

   Both commands return a security group ID, which you use in the next step.

2. Add the rule to the security group using one of the following commands:
• authorize-security-group-ingress (AWS CLI)

```bash
aws ec2 authorize-security-group-ingress --group-id security_group_id --protocol tcp --port 3389 --cidr cidr_ip_range
```

• Grant-EC2SecurityGroupIngress (AWS Tools for Windows PowerShell)

The `Grant-EC2SecurityGroupIngress` command needs an `IpPermission` parameter, which describes the protocol, port range, and IP address range to be used for the security group rule. The following command creates the `IpPermission` parameter:

```powershell
PS C:\> $ip1 = @{ IpProtocol="tcp"; FromPort="3389"; ToPort="3389"; IpRanges="cidr_ip_range" }

PS C:\> Grant-EC2SecurityGroupIngress -GroupId security_group_id -IpPermission @($ip1)
```

Assigning a Security Group to an Instance

You can assign a security group to an instance when you launch the instance. When you add or remove rules, those changes are automatically applied to all instances to which you've assigned the security group.

After you launch an instance, you can change its security groups. For more information, see Changing an Instance's Security Groups in the Amazon VPC User Guide.

Amazon EC2 Instance IP Addressing

Amazon EC2 and Amazon VPC support both the IPv4 and IPv6 addressing protocols. By default, Amazon EC2 and Amazon VPC use the IPv4 addressing protocol; you can't disable this behavior. When you create a VPC, you must specify an IPv4 CIDR block (a range of private IPv4 addresses). You can optionally assign an IPv6 CIDR block to your VPC and subnets, and assign IPv6 addresses from that block to instances in your subnet. IPv6 addresses are reachable over the Internet. For more information about IPv6, see IP Addressing in Your VPC in the Amazon VPC User Guide.

Contents

- Private IPv4 Addresses and Internal DNS Hostnames (p. 656)
- Public IPv4 Addresses and External DNS Hostnames (p. 657)
- Elastic IP Addresses (IPv4) (p. 658)
- Amazon DNS Server (p. 658)
- IPv6 Addresses (p. 658)
- Working with IP Addresses for Your Instance (p. 659)
- Multiple IP Addresses (p. 663)

Private IPv4 Addresses and Internal DNS Hostnames

A private IPv4 address is an IP address that's not reachable over the Internet. You can use private IPv4 addresses for communication between instances in the same VPC. For more information about the
standards and specifications of private IPv4 addresses, see RFC 1918. We allocate private IPv4 addresses to instances using DHCP.

Note
You can create a VPC with a publicly routable CIDR block that falls outside of the private IPv4 address ranges specified in RFC 1918. However, for the purposes of this documentation, we refer to private IPv4 addresses (or ‘private IP addresses’) as the IP addresses that are within the IPv4 CIDR range of your VPC.

When you launch an instance, we allocate a primary private IPv4 address for the instance. Each instance is also given an internal DNS hostname that resolves to the primary private IPv4 address; for example, ip-10-251-50-12.ec2.internal. You can use the internal DNS hostname for communication between instances in the same VPC, but we can't resolve the internal DNS hostname outside of the VPC.

An instance receives a primary private IP address from the IPv4 address range of the subnet. For more information, see VPC and Subnet Sizing in the Amazon VPC User Guide. If you don’t specify a primary private IP address when you launch the instance, we select an available IP address in the subnet's IPv4 range for you. Each instance has a default network interface (eth0) that is assigned the primary private IPv4 address. You can also specify additional private IPv4 addresses, known as secondary private IPv4 addresses. Unlike primary private IP addresses, secondary private IP addresses can be reassigned from one instance to another. For more information, see Multiple IP Addresses (p. 663).

A private IPv4 address remains associated with the network interface when the instance is stopped and restarted, and is released when the instance is terminated.

Public IPv4 Addresses and External DNS Hostnames

A public IP address is an IPv4 address that's reachable from the Internet. You can use public addresses for communication between your instances and the Internet.

Each instance that receives a public IP address is also given an external DNS hostname; for example, ec2-203-0-113-25.compute-1.amazonaws.com. We resolve an external DNS hostname to the public IP address of the instance from outside its VPC, and to the private IPv4 address of the instance from inside its VPC. The public IP address is mapped to the primary private IP address through network address translation (NAT). For more information, see RFC 1631: The IP Network Address Translator (NAT).

When you launch an instance in a default VPC, we assign it a public IP address by default. When you launch an instance into a nondefault VPC, the subnet has an attribute that determines whether instances launched into that subnet receive a public IP address from the public IPv4 address pool. By default, we don’t assign a public IP address to instances launched in a nondefault subnet.

You can control whether your instance receives a public IP address as follows:

- Modifying the public IP addressing attribute of your subnet. For more information, see Modifying the Public IPv4 Addressing Attribute for Your Subnet in the Amazon VPC User Guide.
- Enabling or disabling the public IP addressing feature during launch, which overrides the subnet's public IP addressing attribute. For more information, see Assigning a Public IPv4 Address During Instance Launch (p. 661).

A public IP address is assigned to your instance from Amazon's pool of public IPv4 addresses, and is not associated with your AWS account. When a public IP address is disassociated from your instance, it is released back into the public IPv4 address pool, and you cannot reuse it.

You cannot manually associate or disassociate a public IP address from your instance. Instead, in certain cases, we release the public IP address from your instance, or assign it a new one:

- We release your instance's public IP address when it is stopped or terminated. Your stopped instance receives a new public IP address when it is restarted.
• We release your instance's public IP address when you associate an Elastic IP address with it. When you
disassociate the Elastic IP address from your instance, it receives a new public IP address.
• If the public IP address of your instance in a VPC has been released, it will not receive a new one if
there is more than one network interface attached to your instance.
• If your instance's public IP address is released while it has a secondary private IP address that is
associated with an Elastic IP address, the instance does not receive a new public IP address.

If you require a persistent public IP address that can be associated to and from instances as you require,
use an Elastic IP address instead.

If you use dynamic DNS to map an existing DNS name to a new instance's public IP address, it might take
up to 24 hours for the IP address to propagate through the Internet. As a result, new instances might
not receive traffic while terminated instances continue to receive requests. To solve this problem, use an
Elastic IP address. You can allocate your own Elastic IP address, and associate it with your instance. For
more information, see Elastic IP Addresses (p. 674).

If you assign an Elastic IP address to an instance, it receives an IPv4 DNS hostname if DNS hostnames are
enabled. For more information, see Using DNS with Your VPC in the Amazon VPC User Guide.

Note
Instances that access other instances through their public NAT IP address are charged for
regional or Internet data transfer, depending on whether the instances are in the same region.

Elastic IP Addresses (IPv4)

An Elastic IP address is a public IPv4 address that you can allocate to your account. You can associate it
to and from instances as you require, and it's allocated to your account until you choose to release it. For
more information about Elastic IP addresses and how to use them, see Elastic IP Addresses (p. 674).

We do not support Elastic IP addresses for IPv6.

Amazon DNS Server

Amazon provides a DNS server that resolves Amazon-provided IPv4 DNS hostnames to IPv4 addresses.
The Amazon DNS server is located at the base of your VPC network range plus two. For more
information, see Amazon DNS Server in the Amazon VPC User Guide.

IPv6 Addresses

You can optionally associate an IPv6 CIDR block with your VPC, and associate IPv6 CIDR blocks with
your subnets. The IPv6 CIDR block for your VPC is automatically assigned from Amazon's pool of IPv6
addresses; you cannot choose the range yourself. For more information, see the following topics in the
Amazon VPC User Guide:

• VPC and Subnet Sizing for IPv6
• Associating an IPv6 CIDR Block with Your VPC
• Associating an IPv6 CIDR Block with Your Subnet

IPv6 addresses are globally unique, and therefore reachable over the Internet. Your instance receives an
IPv6 address if an IPv6 CIDR block is associated with your VPC and subnet, and if one of the following is
true:

• Your subnet is configured to automatically assign an IPv6 address to an instance during launch. For
more information, see Modifying the IPv6 Addressing Attribute for Your Subnet.
You assign an IPv6 address to your instance during launch.
- You assign an IPv6 address to the primary network interface of your instance after launch.
- You assign an IPv6 address to a network interface in the same subnet, and attach the network interface to your instance after launch.

When your instance receives an IPv6 address during launch, the address is associated with the primary network interface (eth0) of the instance. You can disassociate the IPv6 address from the network interface. We do not support IPv6 DNS hostnames for your instance.

An IPv6 address persists when you stop and start your instance, and is released when you terminate your instance. You cannot reassign an IPv6 address while it's assigned to another network interface—you must first unassign it.

You can assign additional IPv6 addresses to your instance by assigning them to a network interface attached to your instance. The number of IPv6 addresses you can assign to a network interface and the number of network interfaces you can attach to an instance varies per instance type. For more information, see IP Addresses Per Network Interface Per Instance Type (p. 681).

### Working with IP Addresses for Your Instance

You can view the IP addresses assigned to your instance, assign a public IPv4 address to your instance during launch, or assign an IPv6 address to your instance during launch.

#### Contents
- Determining Your Public, Private, and Elastic IP Addresses (p. 659)
- Determining Your IPv6 Addresses (p. 660)
- Assigning a Public IPv4 Address During Instance Launch (p. 661)
- Assigning an IPv6 Address to an Instance (p. 662)
- Unassigning an IPv6 Address From an Instance (p. 662)

#### Determining Your Public, Private, and Elastic IP Addresses

You can use the Amazon EC2 console to determine the private IPv4 addresses, public IPv4 addresses, and Elastic IP addresses of your instances. You can also determine the public IPv4 and private IPv4 addresses of your instance from within your instance by using instance metadata. For more information, see Instance Metadata and User Data (p. 478).

**To determine your instance's private IPv4 addresses using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**.
3. Select your instance. In the details pane, get the private IPv4 address from the **Private IPs** field, and get the internal DNS hostname from the **Private DNS** field.
4. If you have one or more secondary private IPv4 addresses assigned to network interfaces that are attached to your instance, get those IP addresses from the **Secondary private IPs** field.
5. Alternatively, in the navigation pane, choose **Network Interfaces**, and then select the network interface that's associated with your instance.
6. Get the primary private IP address from the **Primary private IPv4 IP** field, and the internal DNS hostname from the **Private DNS (IPv4)** field.
7. If you've assigned secondary private IP addresses to the network interface, get those IP addresses from the **Secondary private IPv4 IPs** field.
To determine your instance’s public IPv4 addresses using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select your instance. In the details pane, get the public IP address from the IPv4 Public IP field, and get the external DNS hostname from the Public DNS (IPv4) field.
4. If one or more Elastic IP addresses have been associated with the instance, get the Elastic IP addresses from the Elastic IPs field.

**Note**
If your instance does not have a public IPv4 address, but you’ve associated an Elastic IP address with a network interface for the instance, the IPv4 Public IP field displays the Elastic IP address.

5. Alternatively, in the navigation pane, choose Network Interfaces, and then select a network interface that’s associated with your instance.
6. Get the public IP address from the IPv4 Public IP field. An asterisk (*) indicates the public IPv4 address or Elastic IP address that’s mapped to the primary private IPv4 address.

**Note**
The public IPv4 address is displayed as a property of the network interface in the console, but it’s mapped to the primary private IPv4 address through NAT. Therefore, if you inspect the properties of your network interface on your instance, for example, through `ifconfig` (Linux) or `ipconfig` (Windows), the public IPv4 address is not displayed. To determine your instance’s public IPv4 address from within the instance, you can use instance metadata.

To determine your instance’s IPv4 addresses using instance metadata

1. Connect to your instance.
2. Use the following command to access the private IP address:

   ```powershell
   ```

3. Use the following command to access the public IP address:

   ```powershell
   ```

   Note that if an Elastic IP address is associated with the instance, the value returned is that of the Elastic IP address.

Determining Your IPv6 Addresses

You can use the Amazon EC2 console to determine the IPv6 addresses of your instances.

To determine your instance’s IPv6 addresses using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select your instance. In the details pane, get the IPv6 addresses from IPv6 IPs.

To determine your instance’s IPv6 addresses using instance metadata

1. Connect to your instance.
2. Use the following command to view the IPv6 address (you can get the MAC address from http://169.254.169.254/latest/meta-data/network/interfaces/macs/):

```
```

## Assigning a Public IPv4 Address During Instance Launch

Each subnet has an attribute that determines whether instances launched into that subnet are assigned a public IP address. By default, nondefault subnets have this attribute set to false, and default subnets have this attribute set to true. When you launch an instance, a public IPv4 addressing feature is also available for you to control whether your instance is assigned a public IPv4 address; you can override the default behavior of the subnet's IP addressing attribute. The public IPv4 address is assigned from Amazon's pool of public IPv4 addresses, and is assigned to the network interface with the device index of eth0. This feature depends on certain conditions at the time you launch your instance.

**Important**

You can't manually disassociate the public IP address from your instance after launch. Instead, it's automatically released in certain cases, after which you cannot reuse it. For more information, see Public IPv4 Addresses and External DNS Hostnames (p. 657). If you require a persistent public IP address that you can associate or disassociate at will, assign an Elastic IP address to the instance after launch instead. For more information, see Elastic IP Addresses (p. 674).

### To access the public IP addressing feature when launching an instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose **Launch Instance**.
3. Select an AMI and an instance type, and then choose **Next: Configure Instance Details**.
4. On the **Configure Instance Details** page, for **Network**, select a VPC. The **Auto-assign Public IP** list is displayed. Choose **Enable** or **Disable** to override the default setting for the subnet.

**Important**

You cannot auto-assign a public IP address if you specify more than one network interface. Additionally, you cannot override the subnet setting using the auto-assign public IP feature if you specify an existing network interface for eth0.

5. Follow the steps on the next pages of the wizard to complete your instance's setup. For more information about the wizard configuration options, see Launching an Instance Using the Launch Instance Wizard (p. 334). On the final **Review Instance Launch** page, review your settings, and then choose **Launch** to choose a key pair and launch your instance.

6. On the **Instances** page, select your new instance and view its public IP address in **IPv4 Public IP** field in the details pane.

The public IP addressing feature is only available during launch. However, whether you assign a public IP address to your instance during launch or not, you can associate an Elastic IP address with your instance after it's launched. For more information, see Elastic IP Addresses (p. 674). You can also modify your subnet's public IPv4 addressing behavior. For more information, see Modifying the Public IPv4 Addressing Attribute for Your Subnet.

### To enable or disable the public IP addressing feature using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- Use the **--associate-public-ip-address** or the **--no-associate-public-ip-address** option with the **run-instances** command (AWS CLI)

661
Assigning an IPv6 Address to an Instance

If your VPC and subnet have IPv6 CIDR blocks associated with them, you can assign an IPv6 address to your instance during or after launch. The IPv6 address is assigned from the IPv6 address range of the subnet, and is assigned to the network interface with the device index of eth0.

IPv6 is supported on all current generation instance types and the C3, R3, and I2 previous generation instance types.

To assign an IPv6 address to an instance during launch

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Select an AMI and an instance type that supports IPv6, and choose Next: Configure Instance Details.
3. On the Configure Instance Details page, for Network, select a VPC and for Subnet, select a subnet. For Auto-assign IPv6 IP, choose Enable.
4. Follow the remaining steps in the wizard to launch your instance.

Alternatively, you can assign an IPv6 address to your instance after launch.

To assign an IPv6 address to your instance after launch

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select your instance, choose Actions, Networking, Manage IP Addresses.
4. Under IPv6 Addresses, choose Assign new IP. You can specify an IPv6 address from the range of the subnet, or leave the Auto-assign value to let Amazon choose an IPv6 address for you.
5. Choose Save.

Note
If you launched your instance using Amazon Linux 2016.09.0 or later, or Windows Server 2008 R2 or later, your instance is configured for IPv6, and no additional steps are needed to ensure that the IPv6 address is recognized on the instance. If you launched your instance from an older AMI, you may have to configure your instance manually. For more information, see Configure IPv6 on Your Instances in the Amazon VPC User Guide.

To assign an IPv6 address using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- Use the --ipv6-addresses option with the run-instances command (AWS CLI)
- Use the Ipv6Addresses property for -NetworkInterface in the New-EC2Instance command (AWS Tools for Windows PowerShell)
- assign-ipv6-addresses (AWS CLI)
- Register-EC2Ipv6AddressList (AWS Tools for Windows PowerShell)

Unassigning an IPv6 Address From an Instance

You can unassign an IPv6 address from an instance using the Amazon EC2 console.
To unassign an IPv6 address from an instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select your instance, choose Actions, Networking, Manage IP Addresses.
4. Under IPv6 Addresses, choose Unassign for the IPv6 address to unassign.
5. Choose Yes, Update.

To unassign an IPv6 address using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- unassign-ipv6-addresses (AWS CLI)

Multiple IP Addresses

You can specify multiple private IPv4 and IPv6 addresses for your instances. The number of network interfaces and private IPv4 and IPv6 addresses that you can specify for an instance depends on the instance type. For more information, see IP Addresses Per Network Interface Per Instance Type (p. 681).

It can be useful to assign multiple IP addresses to an instance in your VPC to do the following:

- Host multiple websites on a single server by using multiple SSL certificates on a single server and associating each certificate with a specific IP address.
- Operate network appliances, such as firewalls or load balancers, that have multiple IP addresses for each network interface.
- Redirect internal traffic to a standby instance in case your instance fails, by reassigning the secondary IP address to the standby instance.

Contents

- How Multiple IP Addresses Work (p. 663)
- Working with Multiple IPv4 Addresses (p. 664)
- Working with Multiple IPv6 Addresses (p. 667)

How Multiple IP Addresses Work

The following list explains how multiple IP addresses work with network interfaces:

- You can assign a secondary private IPv4 address to any network interface. The network interface need not be attached to the instance.
- You can assign multiple IPv6 addresses to a network interface that's in a subnet that has an associated IPv6 CIDR block.
- You must choose a secondary IPv4 address from the IPv4 CIDR block range of the subnet for the network interface.
- You must choose IPv6 addresses from the IPv6 CIDR block range of the subnet for the network interface.
- You associate security groups with network interfaces, not individual IP addresses. Therefore, each IP address you specify in a network interface is subject to the security group of its network interface.
Multiple IP addresses can be assigned and unassigned to network interfaces attached to running or stopped instances.

Secondary private IPv4 addresses that are assigned to a network interface can be reassigned to another one if you explicitly allow it.

An IPv6 address cannot be reassigned to another network interface; you must first unassign the IPv6 address from the existing network interface.

When assigning multiple IP addresses to a network interface using the command line tools or API, the entire operation fails if one of the IP addresses can't be assigned.

Primary private IPv4 addresses, secondary private IPv4 addresses, Elastic IP addresses, and IPv6 addresses remain with a secondary network interface when it is detached from an instance or attached to an instance.

Although you can't detach the primary network interface from an instance, you can reassign the secondary private IPv4 address of the primary network interface to another network interface.

The following list explains how multiple IP addresses work with Elastic IP addresses (IPv4 only):

- Each private IPv4 address can be associated with a single Elastic IP address, and vice versa.
- When a secondary private IPv4 address is reassigned to another interface, the secondary private IPv4 address retains its association with an Elastic IP address.
- When a secondary private IPv4 address is unassigned from an interface, an associated Elastic IP address is automatically disassociated from the secondary private IPv4 address.

### Working with Multiple IPv4 Addresses

You can assign a secondary private IPv4 address to an instance, associate an Elastic IPv4 address with a secondary private IPv4 address, and unassign a secondary private IPv4 address.

#### Contents

- Assigning a Secondary Private IPv4 Address (p. 664)
- Configuring the Operating System on Your Instance to Recognize the Secondary Private IPv4 Address (p. 666)
- Associating an Elastic IP Address with the Secondary Private IPv4 Address (p. 666)
- Viewing Your Secondary Private IPv4 Addresses (p. 666)
- Unassigning a Secondary Private IPv4 Address (p. 667)

#### Assigning a Secondary Private IPv4 Address

You can assign the secondary private IPv4 address to the network interface for an instance as you launch the instance, or after the instance is running. This section includes the following procedures.

- To assign a secondary private IPv4 address when launching an instance (p. 664)
- To assign a secondary IPv4 address during launch using the command line (p. 665)
- To assign a secondary private IPv4 address to a network interface (p. 665)
- To assign a secondary private IPv4 to an existing instance using the command line (p. 666)

#### To assign a secondary private IPv4 address when launching an instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose **Launch Instance**.
3. Select an AMI, then choose an instance type and choose Next: Configure Instance Details.
4. On the Configure Instance Details page, for Network, select a VPC and for Subnet, select a subnet.
5. In the Network Interfaces section, do the following, and then choose Next: Add Storage:
   - To add another network interface, choose Add Device. The console enables you to specify up to two network interfaces when you launch an instance. After you launch the instance, choose Network Interfaces in the navigation pane to add additional network interfaces. The total number of network interfaces that you can attach varies by instance type. For more information, see IP Addresses Per Network Interface Per Instance Type (p. 681).
     
     Important
     When you add a second network interface, the system can no longer auto-assign a public IPv4 address. You will not be able to connect to the instance over IPv4 unless you assign an Elastic IP address to the primary network interface (eth0). You can assign the Elastic IP address after you complete the Launch wizard. For more information, see Working with Elastic IP Addresses (p. 675).
     - For each network interface, under Secondary IP addresses, choose Add IP, and then enter a private IP address from the subnet range, or accept the default Auto-assign value to let Amazon select an address.
6. On the next Add Storage page, you can specify volumes to attach to the instance besides the volumes specified by the AMI (such as the root device volume), and then choose Next: Add Tags.
7. On the Add Tags page, specify tags for the instance, such as a user-friendly name, and then choose Next: Configure Security Group.
8. On the Configure Security Group page, select an existing security group or create a new one. Choose Review and Launch.
9. On the Review Instance Launch page, review your settings, and then choose Launch to choose a key pair and launch your instance. If you're new to Amazon EC2 and haven't created any key pairs, the wizard prompts you to create one.

     Important
     After you have added a secondary private IP address to a network interface, you must connect to the instance and configure the secondary private IP address on the instance itself. For more information, see Configuring the Operating System on Your Instance to Recognize the Secondary Private IPv4 Address (p. 666).

To assign a secondary IPv4 address during launch using the command line

- You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
  - The --secondary-private-ip-addresses option with the run-instances command (AWS CLI)
  - Define -NetworkInterface and specify the PrivateIpAddresses parameter with the New-EC2Instance command (AWS Tools for Windows PowerShell).

To assign a secondary private IPv4 address to a network interface

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces, and then select the network interface attached to the instance.
3. Choose Actions, Manage IP Addresses.
4. Under IPv4 Addresses, choose Assign new IP.
5. Enter a specific IPv4 address that's within the subnet range for the instance, or leave the field blank to let Amazon select an IP address for you.
6. (Optional) Choose **Allow reassignment** to allow the secondary private IP address to be reassigned if it is already assigned to another network interface.
7. Choose **Yes, Update**.

Alternatively, you can assign a secondary private IPv4 address to an instance. Choose **Instances** in the navigation pane, select the instance, and then choose **Actions, Networking, Manage IP Addresses**. You can configure the same information as you did in the steps above. The IP address is assigned to the primary network interface (eth0) for the instance.

**To assign a secondary private IPv4 to an existing instance using the command line**

- You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).
  - `assign-private-ip-addresses` (AWS CLI)
  - `Register-EC2PrivatelpAddress` (AWS Tools for Windows PowerShell)

### Configuring the Operating System on Your Instance to Recognize the Secondary Private IPv4 Address

After you assign a secondary private IPv4 address to your instance, you need to configure the operating system on your instance to recognize the secondary private IP address.

For information about configuring a Windows instance, see **Configuring a Secondary Private IPv4 Address for Your Windows Instance** (p. 468).

### Associating an Elastic IP Address with the Secondary Private IPv4 Address

**To associate an Elastic IP address with a secondary private IPv4 address**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Elastic IPs**.
3. Choose **Actions**, and then select **Associate address**.
4. For **Network interface**, select the network interface, and then select the secondary IP address from the **Private IP** list.
5. Choose **Associate**.

**To associate an Elastic IP address with a secondary private IPv4 address using the command line**

- You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).
  - `associate-address` (AWS CLI)
  - `Register-EC2Address` (AWS Tools for Windows PowerShell)

### Viewing Your Secondary Private IPv4 Addresses

**To view the private IPv4 addresses assigned to a network interface**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Network Interfaces**.
3. Select the network interface with private IP addresses to view.
4. On the Details tab in the details pane, check the **Primary private IPv4 IP** and **Secondary private IPv4 IPs** fields for the primary private IPv4 address and any secondary private IPv4 addresses assigned to the network interface.

**To view the private IPv4 addresses assigned to an instance**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**.
3. Select the instance with private IPv4 addresses to view.
4. On the **Description** tab in the details pane, check the **Private IPs** and **Secondary private IPs** fields for the primary private IPv4 address and any secondary private IPv4 addresses assigned to the instance through its network interface.

**Unassigning a Secondary Private IPv4 Address**

If you no longer require a secondary private IPv4 address, you can unassign it from the instance or the network interface. When a secondary private IPv4 address is unassigned from a network interface, the Elastic IP address (if it exists) is also disassociated.

**To unassign a secondary private IPv4 address from an instance**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**.
3. Select an instance, choose **Actions, Networking, Manage IP Addresses**.
4. Under **IPv4 Addresses**, choose **Unassign** for the IPv4 address to unassign.
5. Choose **Yes, Update**.

**To unassign a secondary private IPv4 address from a network interface**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Network Interfaces**.
3. Select the network interface, choose **Actions, Manage IP Addresses**.
4. Under **IPv4 Addresses**, choose **Unassign** for the IPv4 address to unassign.
5. Choose **Yes, Update**.

**To unassign a secondary private IPv4 address using the command line**

- You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
  - `unassign-private-ip-addresses` (AWS CLI)
  - `Unregister-EC2PrivateIpAddress` (AWS Tools for Windows PowerShell)

**Working with Multiple IPv6 Addresses**

You can assign multiple IPv6 addresses to your instance, view the IPv6 addresses assigned to your instance, and unassign IPv6 addresses from your instance.
Assigning Multiple IPv6 Addresses

You can assign one or more IPv6 addresses to your instance during launch or after launch. To assign an IPv6 address to an instance, the VPC and subnet in which you launch the instance must have an associated IPv6 CIDR block. For more information, see VPCs and Subnets in the Amazon VPC User Guide.

To assign multiple IPv6 addresses during launch

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the dashboard, choose Launch Instance.
3. Select an AMI, choose an instance type, and choose Next: Configure Instance Details. Ensure that you choose an instance type that support IPv6. For more information, see Instance Types (p. 119).
4. On the Configure Instance Details page, select a VPC from the Network list, and a subnet from the Subnet list.
5. In the Network Interfaces section, do the following, and then choose Next: Add Storage:
   - To assign a single IPv6 address to the primary network interface (eth0), under IPv6 IPs, choose Add IP. To add a secondary IPv6 address, choose Add IP again. You can enter an IPv6 address from the range of the subnet, or leave the default Auto-assigned value to let Amazon choose an IPv6 address from the subnet for you.
   - Choose Add Device to add another network interface and repeat the steps above to add one or more IPv6 addresses to the network interface. The console enables you to specify up to two network interfaces when you launch an instance. After you launch the instance, choose Network Interfaces in the navigation pane to add additional network interfaces. The total number of network interfaces that you can attach varies by instance type. For more information, see IP Addresses Per Network Interface Per Instance Type (p. 681).
6. Follow the next steps in the wizard to attach volumes and tag your instance.
7. On the Configure Security Group page, select an existing security group or create a new one. If you want your instance to be reachable over IPv6, ensure that your security group has rules that allow access from IPv6 addresses. For more information, see Security Group Rules Reference (p. 591). Choose Review and Launch.
8. On the Review Instance Launch page, review your settings, and then choose Launch to choose a key pair and launch your instance. If you're new to Amazon EC2 and haven't created any key pairs, the wizard prompts you to create one.

You can use the Instances screen Amazon EC2 console to assign multiple IPv6 addresses to an existing instance. This assigns the IPv6 addresses to the primary network interface (eth0) for the instance. To assign a specific IPv6 address to the instance, ensure that the IPv6 address is not already assigned to another instance or network interface.

To assign multiple IPv6 addresses to an existing instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select your instance, choose Actions, Networking, Manage IP Addresses.
4. Under IPv6 Addresses, choose Assign new IP for each IPv6 address you want to add. You can specify an IPv6 address from the range of the subnet, or leave the Auto-assigned value to let Amazon choose an IPv6 address for you.
5. Choose Yes, Update.
Alternatively, you can assign multiple IPv6 addresses to an existing network interface. The network interface must have been created in a subnet that has an associated IPv6 CIDR block. To assign a specific IPv6 address to the network interface, ensure that the IPv6 address is not already assigned to another network interface.

**To assign multiple IPv6 addresses to a network interface**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Network Interfaces**.
4. Under **IPv6 Addresses**, choose **Assign new IP** for each IPv6 address you want to add. You can specify an IPv6 address from the range of the subnet, or leave the **Auto-assign** value to let Amazon choose an IPv6 address for you.
5. Choose **Yes, Update**.

**CLI Overview**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- **Assign an IPv6 address during launch:**
  - Use the **--ipv6-addresses** or **--ipv6-address-count** options with the **run-instances** command (AWS CLI)
  - Define **-NetworkInterface** and specify the **Ipv6Addresses** or **Ipv6AddressCount** parameters with the **New-EC2Instance** command (AWS Tools for Windows PowerShell).
- **Assign an IPv6 address to a network interface:**
  - **assign-ipv6-addresses** (AWS CLI)
  - **Register-EC2ipv6AddressList** (AWS Tools for Windows PowerShell)

**Viewing Your IPv6 Addresses**

You can view the IPv6 addresses for an instance or for a network interface.

**To view the IPv6 addresses assigned to an instance**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Select your instance. In the details pane, review the **IPv6 IPs** field.

**To view the IPv6 addresses assigned to a network interface**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Network Interfaces**.
3. Select your network interface. In the details pane, review the **IPv6 IPs** field.

**CLI Overview**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- **View the IPv6 addresses for an instance:**
  - **describe-instances** (AWS CLI)
• View the IPv6 addresses for a network interface:
  • describe-network-interfaces (AWS CLI)
  • Get-EC2NetworkInterface (AWS Tools for Windows PowerShell)

Unassigning an IPv6 Address

You can unassign an IPv6 address from the primary network interface of an instance, or you can unassign an IPv6 address from a network interface.

To unassign an IPv6 address from an instance
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select your instance, choose Actions, Networking, Manage IP Addresses.
4. Under IPv6 Addresses, choose Unassign for the IPv6 address to unassign.
5. Choose Yes, Update.

To unassign an IPv6 address from a network interface
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Select your network interface, choose Actions, Manage IP Addresses.
4. Under IPv6 Addresses, choose Unassign for the IPv6 address to unassign.
5. Choose Save.

CLI Overview

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

• unassign-ipv6-addresses (AWS CLI)
• Unregister-EC2Ipv6AddressList (AWS Tools for Windows PowerShell).

Bring Your Own IP Addresses (BYOIP)

You can bring part or all of your public IPv4 address range from your on-premises network to your AWS account. You continue to own the address range, but AWS advertises it on the internet. After you bring the address range to AWS, it appears in your account as an address pool. You can create an Elastic IP address from your address pool and use it with your AWS resources, such as EC2 instances, NAT gateways, and Network Load Balancers.

Important
BYOIP is not available in all Regions. For a list of supported Regions, see the FAQ for Bring Your Own IP.

Requirements

• The address range must be registered with your regional internet registry (RIR), such as the American Registry for Internet Numbers (ARIN) or Réseaux IP Européens Network Coordination Centre (RIPE).
It must be registered to a business or institutional entity and may not be registered to an individual person.

- The most specific address range that you can specify is /24.
- You can bring each address range to one region at a time.
- You can bring 5 address ranges per region to your AWS account.
- The addresses in the IP address range must have a clean history. We may investigate the reputation of the IP address range and reserve the right to reject an IP address range if it contains an IP address that has poor reputation or is associated with malicious behavior.
- The following are supported:
  - ARIN - "Direct Allocation" and "Direct Assignment" network types
  - RIPE - "ALLOCATED PA", "LEGACY", and "ASSIGNED PI" allocation statuses

### Prepare to Bring Your Address Range to Your AWS Account

To ensure that only you can bring your address range to your AWS account, you must authorize Amazon to advertise the address range and provide proof that you own the address range through a signed authorization message.

A Route Origin Authorization (ROA) is a document that you can create through your RIR. It contains the address range, the Autonomous System numbers (ASN) that are allowed to advertise the address range, and an expiration date. An ROA authorizes Amazon to advertise an address range under a specific AS number. However, it does not authorize your AWS account to bring the address range to AWS. To authorize your AWS account to bring an address range to AWS, you must publish a self-signed X509 certificate in the Registry Data Access Protocol (RDAP) remarks for the address range. The certificate contains a public key, which AWS uses to verify the authorization-context signature that you provide. You should keep your private key secure and use it to sign the authorization-context message.

The commands in these tasks are supported on Linux. On Windows, you can use the Windows Subsystem for Linux to run Linux commands.

**Tasks**

- Create a ROA Object (p. 671)
- Create a Self-Signed X509 Certificate (p. 671)
- Create a Signed Authorization Message (p. 672)

### Create a ROA Object

Create a ROA object to authorize Amazon ASNs 16509 and 14618 to advertise your address range, plus the ASNs that are currently authorized to advertise the address range. You must set the maximum length to the size of the smallest prefix that you want to bring (for example, /24). It might take up to 24 hours for the ROA to become available to Amazon. For more information, see the following:

- ARIN — ROA Requests
- RIPE — Managing ROAs

### Create a Self-Signed X509 Certificate

Use the following procedure to create a self-signed X509 certificate and add it to the RDAP record for your RIR. The openssl commands require OpenSSL version 1.0.2 or later.
To create a self-signed X509 certificate and add it to the RDAP record

1. Generate an RSA 2048-bit key pair as follows:

   ```bash
   openssl genrsa -out private.key 2048
   ```

2. Create a public X509 certificate from the key pair using the following command. In this example, the certificate expires in 365 days, after which time it cannot be trusted. Therefore, be sure to set the expiration appropriately. When prompted for information, you can accept the default values.

   ```bash
   openssl req -new -x509 -key private.key -days 365 | tr -d "\n" > publickey.cer
   ```

3. Update the RDAP record for your RIR with the X509 certificate. Be sure to copy the `-----BEGIN CERTIFICATE-----` and `-----END CERTIFICATE-----` from the certificate. Be sure that you have removed newline characters, if you haven't already done so using the `tr -d "\n"` commands in the previous steps. To view your certificate, run the following command:

   ```bash
   cat publickey.cer
   ```

   For ARIN, add the certificate in the "Public Comments" section for your address range.

   For RIPE, add the certificate as a new "descr" field for your address range.

Create a Signed Authorization Message

The format of the signed authorization message is as follows, where the date is the expiry date of the message:

```
1|aws|account|cidr|YYYYMMDD|SHA256|RSAPSS
```

First, create a plain-text authorization message and store it in a variable named `text_message` as follows. Replace the example account number, address range, and expiry date with your own values.

```
text_message="1|aws|123456789012|198.51.100.0/24|20191201|SHA256|RSAPSS"
```

Next, sign the authorization message in `text_message` using the key pair that you created and stores it in a variable named `signed_message` as follows.

```
signed_message=$(echo $text_message | tr -d "\n" | openssl dgst -sha256 -sigopt rsa_padding_mode:pss -sigopt rsa_pss_saltlen:-1 -sign private.key -keyform PEM | openssl base64 | tr -- '+=/' '-_~' | tr -d "\n")
```

Provision the Address Range for use with AWS

When you provision an address range for use with AWS, you are confirming that you own the address range and authorizing Amazon to advertise it. We also verify that you own the address range through a signed authorization message. This message is signed with the self-signed X509 key pair you used when updating the RDAP record with the X509 certificate.

To provision the address range, use the following `provision-byoip-cidr` command. Replace the example address range with your own address range. The `--cidr-authorization-context` option uses the variables that you created previously, not the ROA message.
Provisioning an address range is an asynchronous operation, so the call returns immediately, but the address range is not ready to use until its status changes from pending-provision to provisioned. It can take up to five days to complete the provisioning process. To monitor the status of the address ranges that you’ve provisioned, use the following describe-byoip-cidrs command:

```
aws ec2 describe-byoip-cidrs --max-results 5
```

To create an Elastic IP address from your address pool, use the allocate-address command. You can use the --public-ipv4-pool option to specify the ID of the address pool returned by describe-byoip-cidrs, or use the --address option to specify an address from the address range that you provisioned.

### Advertise the Address Range through AWS

After the address range is provisioned, it is ready to be advertised. You must advertise the exact address range that you provisioned. You can't advertise only a portion of the provisioned address range.

We recommend that you stop advertising the address range from other locations before you advertise it through AWS. If you keep advertising your IP address range from other locations, we can't reliably support it or troubleshoot issues. Specifically, we can't guarantee that traffic to the address range will enter our network.

To minimize down time, you can configure your AWS resources to use an address from your address pool before it is advertised, and then simultaneously stop advertising it from the current location and start advertising it through AWS. For more information about allocating an Elastic IP address from your address pool, see Allocating an Elastic IP Address (p. 675).

To advertise the address range, use the following advertise-byoip-cidr command:

```
aws ec2 advertise-byoip-cidr --cidr address-range
```

**Important**

You can run the advertise-byoip-cidr command at most once every 10 seconds, even if you specify different address ranges each time.

To stop advertising the address range, use the following withdraw-byoip-cidr command:

```
aws ec2 withdraw-byoip-cidr --cidr address-range
```

**Important**

You can run the withdraw-byoip-cidr command at most once every 10 seconds, even if you specify different address ranges each time.

### Deprovision the Address Range

To stop using your address range with AWS, release any Elastic IP addresses still allocated from the address pool, stop advertising the address range, and deprovision the address range.

To release each Elastic IP address, use the following release-address command:

```
aws ec2 release-address --allocation-id eipalloc-12345678
```
Elastic IP Addresses

An Elastic IP address is a static IPv4 address designed for dynamic cloud computing. An Elastic IP address is associated with your AWS account. With an Elastic IP address, you can mask the failure of an instance or software by rapidly remapping the address to another instance in your account.

An Elastic IP address is a public IPv4 address, which is reachable from the internet. If your instance does not have a public IPv4 address, you can associate an Elastic IP address with your instance to enable communication with the internet; for example, to connect to your instance from your local computer.

We currently do not support Elastic IP addresses for IPv6.

Contents

• Elastic IP Address Basics (p. 674)
• Working with Elastic IP Addresses (p. 675)
• Using Reverse DNS for Email Applications (p. 679)
• Elastic IP Address Limit (p. 679)

Elastic IP Address Basics

The following are the basic characteristics of an Elastic IP address:

• To use an Elastic IP address, you first allocate one to your account, and then associate it with your instance or a network interface.

• When you associate an Elastic IP address with an instance or its primary network interface, the instance's public IPv4 address (if it had one) is released back into Amazon's pool of public IPv4 addresses. You cannot reuse a public IPv4 address, and you cannot convert a public IPv4 address to an Elastic IP address. For more information, see Public IPv4 Addresses and External DNS Hostnames (p. 657).

• You can disassociate an Elastic IP address from a resource, and reassociate it with a different resource. Any open connections to an instance continue to work for a time even after you disassociate its Elastic IP address and reassociate it with another instance. We recommend that you reopen these connections using the reassociated Elastic IP address.

• A disassociated Elastic IP address remains allocated to your account until you explicitly release it.

• To ensure efficient use of Elastic IP addresses, we impose a small hourly charge if an Elastic IP address is not associated with a running instance, or if it is associated with a stopped instance or an unattached network interface. While your instance is running, you are not charged for one Elastic IP address associated with the instance, but you are charged for any additional Elastic IP addresses associated with the instance. For more information, see Amazon EC2 Pricing.

• An Elastic IP address is for use in a specific Region only.

• When you associate an Elastic IP address with an instance that previously had a public IPv4 address, the public DNS hostname of the instance changes to match the Elastic IP address.
• We resolve a public DNS hostname to the public IPv4 address or the Elastic IP address of the instance outside the network of the instance, and to the private IPv4 address of the instance from within the network of the instance.

• When you allocate an Elastic IP address from an IP address pool that you have brought to your AWS account, it does not count toward your Elastic IP address limits.

Working with Elastic IP Addresses

The following sections describe how you can work with Elastic IP addresses.

Tasks
• Allocating an Elastic IP Address (p. 675)
• Describing Your Elastic IP Addresses (p. 676)
• Tagging an Elastic IP Address (p. 676)
• Associating an Elastic IP Address with a Running Instance (p. 677)
• Disassociating an Elastic IP Address and Reassociating with a Different Instance (p. 677)
• Releasing an Elastic IP Address (p. 678)
• Recovering an Elastic IP Address (p. 678)

Allocating an Elastic IP Address

You can allocate an Elastic IP address from Amazon's pool of public IPv4 addresses, or from a custom IP address pool that you have brought to your AWS account. For more information about bringing your own IP address range to your AWS account, see Bring Your Own IP Addresses (BYOIP) (p. 670).

You can allocate an Elastic IP address using the Amazon EC2 console or the command line.

To allocate an Elastic IP address from Amazon's pool of public IPv4 addresses using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Choose Allocate new address.
4. For IPv4 address pool, choose Amazon pool.
5. Choose Allocate, and close the confirmation screen.

To allocate an Elastic IP address from an IP address pool that you own using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Choose Allocate new address.
4. For IPv4 address pool, choose Owned by me and then select the IP address pool.
   To see the IP address range of the selected address pool and the number of IP addresses already allocated from the address pool, see Address ranges.
5. For IPv4 address, do one of the following:
   • To let Amazon EC2 select an IP address from the address pool, choose No preference.
   • To select a specific IP address from the address pool, choose Select an address and then type the IP address.
6. Choose Allocate, and close the confirmation screen.
To allocate an Elastic IP address using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- allocate-address (AWS CLI)
- New-EC2Address (AWS Tools for Windows PowerShell)

Describing Your Elastic IP Addresses

You can describe an Elastic IP address using the Amazon EC2 or the command line.

To describe your Elastic IP addresses using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Select a filter from the Resource Attribute list to begin searching. You can use multiple filters in a single search.

To describe your Elastic IP addresses using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- describe-addresses (AWS CLI)
- Get-EC2Address (AWS Tools for Windows PowerShell)

Tagging an Elastic IP Address

You can assign custom tags to your Elastic IP addresses to categorize them in different ways, for example, by purpose, owner, or environment. This helps you to quickly find a specific Elastic IP address based on the custom tags you've assigned it.

Note
Cost allocation tracking using Elastic IP address tags is not supported.

To tag an Elastic IP address using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Select the Elastic IP address to tag and choose Tags.
4. Choose Add/Edit Tags.
5. In the Add/Edit Tags dialog box, choose Create Tag, and then specify the key and value for the tag.
6. (Optional) Choose Create Tag to add additional tags to the Elastic IP address.
7. Choose Save.

To tag an Elastic IP address using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- create-tags (AWS CLI)
`aws ec2 create-tags --resources eipalloc-12345678 --tags Key=Owner,Value=TeamA`

- **New-EC2Tag (AWS Tools for Windows PowerShell)**

  The `New-EC2Tag` command needs a `Tag` parameter, which specifies the key and value pair to be used for the Elastic IP address tag. The following commands create the `Tag` parameter:

  ```powershell
  PS C:\> $tag = New-Object Amazon.EC2.Model.Tag
  PS C:\> $tag.Key = "Owner"
  PS C:\> $tag.Value = "TeamA"
  ```

  ```powershell
  PS C:\> New-EC2Tag -Resource eipalloc-12345678 -Tag $tag
  ```

**Associating an Elastic IP Address with a Running Instance**

You can associate an Elastic IP address to an instance using the Amazon EC2 console or the command line.

If you're associating an Elastic IP address with your instance to enable communication with the internet, you must also ensure that your instance is in a public subnet. For more information, see Internet Gateways in the Amazon VPC User Guide.

**To associate an Elastic IP address with an instance using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Select an Elastic IP address and choose Actions, Associate address.
4. Select the instance from Instance and then choose Associate.

**To associate an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- **associate-address (AWS CLI)**
- **Register-EC2Address (AWS Tools for Windows PowerShell)**

**Disassociating an Elastic IP Address and Reassociating with a Different Instance**

You can disassociate an Elastic IP address and then reassociate it using the Amazon EC2 console or the command line.

**To disassociate and reassociate an Elastic IP address using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Select the Elastic IP address, choose Actions, and then select Disassociate address.
4. Choose Disassociate address.
5. Select the address that you disassociated in the previous step. For Actions, choose Associate address.
6. Select the new instance from **Instance**, and then choose **Associate**.

**To disassociate an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2 (p. 3)**.

- `disassociate-address` (AWS CLI)
- `Unregister-EC2Address` (AWS Tools for Windows PowerShell)

**To associate an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2 (p. 3)**.

- `associate-address` (AWS CLI)
- `Register-EC2Address` (AWS Tools for Windows PowerShell)

**Releasing an Elastic IP Address**

If you no longer need an Elastic IP address, we recommend that you release it (the address must not be associated with an instance).

**To release an Elastic IP address using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Elastic IPs**.
3. Select the Elastic IP address, choose **Actions**, and then select **Release addresses**. Choose **Release** when prompted.

**To release an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2 (p. 3)**.

- `release-address` (AWS CLI)
- `Remove-EC2Address` (AWS Tools for Windows PowerShell)

**Recovering an Elastic IP Address**

If you have released your Elastic IP address, you might be able to recover it. The following rules apply:

- You cannot recover an Elastic IP address if it has been allocated to another AWS account, or if it will result in your exceeding your Elastic IP address limit.
- You cannot recover tags associated with an Elastic IP address.
- You can recover an Elastic IP address using the Amazon EC2 API or a command line tool only.

**To recover an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2 (p. 3)**.

- `allocate-address` (AWS CLI) — Specify the IP address using the `--address` parameter as follows.
Using Reverse DNS for Email Applications

If you intend to send email to third parties from an instance, we suggest you provision one or more Elastic IP addresses and provide them to us. AWS works with ISPs and internet anti-spam organizations to reduce the chance that your email sent from these addresses will be flagged as spam.

In addition, assigning a static reverse DNS record to your Elastic IP address used to send email can help avoid having email flagged as spam by some anti-spam organizations. Note that a corresponding forward DNS record (record type A) pointing to your Elastic IP address must exist before we can create your reverse DNS record.

If a reverse DNS record is associated with an Elastic IP address, the Elastic IP address is locked to your account and cannot be released from your account until the record is removed.

To remove email sending limits, or to provide us with your Elastic IP addresses and reverse DNS records, go to the Request to Remove Email Sending Limitations page.

Elastic IP Address Limit

By default, all AWS accounts are limited to five (5) Elastic IP addresses per Region, because public (IPv4) internet addresses are a scarce public resource. We strongly encourage you to use an Elastic IP address primarily for the ability to remap the address to another instance in the case of instance failure, and to use DNS hostnames for all other inter-node communication.

If you feel your architecture warrants additional Elastic IP addresses, you can request a limit increase. To request an increase, complete the Amazon VPC limit request form (choose VPC Elastic IP Address Limit). Describe your use case so that we can understand your needs. To request Elastic IP addresses for EC2-Classic, complete the Amazon EC2 Elastic IP Address limit request form.

Elastic Network Interfaces

An elastic network interface (referred to as a network interface in this documentation) is a logical networking component in a VPC that represents a virtual network card.

A network interface can include the following attributes:

- A primary private IPv4 address from the IPv4 address range of your VPC
- One or more secondary private IPv4 addresses from the IPv4 address range of your VPC
- One Elastic IP address (IPv4) per private IPv4 address
- One public IPv4 address
- One or more IPv6 addresses
- One or more security groups
- A MAC address
- A source/destination check flag
- A description
You can create and configure network interfaces in your account and attach them to instances in your VPC. Your account might also have requester-managed network interfaces, which are created and managed by AWS services to enable you to use other resources and services. You cannot manage these network interfaces yourself. For more information, see Requester-Managed Network Interfaces (p. 699).

All network interfaces have a resource identifier that starts with eni-.

**Important**
The term 'elastic network interface' is sometimes shortened to 'ENI'. This is not the same as the Elastic Network Adapter (ENA), which is a custom interface that optimizes network performance on some instance types. For more information, see Enhanced Networking on Windows (p. 700).

## Network Interface Basics

You can create a network interface, attach it to an instance, detach it from an instance, and attach it to another instance. The attributes of a network interface follow it as it's attached or detached from an instance and reattached to another instance. When you move a network interface from one instance to another, network traffic is redirected to the new instance.

You can also modify the attributes of your network interface, including changing its security groups and managing its IP addresses.

Every instance in a VPC has a default network interface, called the primary network interface (eth0). You cannot detach a primary network interface from an instance. You can create and attach additional network interfaces. The maximum number of network interfaces that you can use varies by instance type. For more information, see IP Addresses Per Network Interface Per Instance Type (p. 681).

### Public IPv4 addresses for network interfaces

In a VPC, all subnets have a modifiable attribute that determines whether network interfaces created in that subnet (and therefore instances launched into that subnet) are assigned a public IPv4 address. For more information, see IP Addressing Behavior for Your Subnet in the Amazon VPC User Guide. The public IPv4 address is assigned from Amazon's pool of public IPv4 addresses. When you launch an instance, the IP address is assigned to the primary network interface (eth0) that's created.

When you create a network interface, it inherits the public IPv4 addressing attribute from the subnet. If you later modify the public IPv4 addressing attribute of the subnet, the network interface keeps the setting that was in effect when it was created. If you launch an instance and specify an existing network interface for eth0, the public IPv4 addressing attribute is determined by the network interface.

For more information, see Public IPv4 Addresses and External DNS Hostnames (p. 657).

### IPv6 addresses for network interfaces

You can associate an IPv6 CIDR block with your VPC and subnet, and assign one or more IPv6 addresses from the subnet range to a network interface.

All subnets have a modifiable attribute that determines whether network interfaces created in that subnet (and therefore instances launched into that subnet) are automatically assigned an IPv6 address.
from the range of the subnet. For more information, see IP Addressing Behavior for Your Subnet in the Amazon VPC User Guide. When you launch an instance, the IPv6 address is assigned to the primary network interface (eth0) that's created.

For more information, see IPv6 Addresses (p. 658).

Monitoring IP Traffic

You can enable a VPC flow log on your network interface to capture information about the IP traffic going to and from a network interface. After you've created a flow log, you can view and retrieve its data in Amazon CloudWatch Logs. For more information, see VPC Flow Logs in the Amazon VPC User Guide.

IP Addresses Per Network Interface Per Instance Type

The following table lists the maximum number of network interfaces per instance type, and the maximum number of private IPv4 addresses and IPv6 addresses per network interface. The limit for IPv6 addresses is separate from the limit for private IPv4 addresses per network interface. Not all instance types support IPv6 addressing. Network interfaces, multiple private IPv4 addresses, and IPv6 addresses are only available for instances running in a VPC. For more information, see Multiple IP Addresses (p. 663). For more information about IPv6 in VPC, see IP Addressing in Your VPC in the Amazon VPC User Guide.

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Maximum Network Interfaces</th>
<th>IPv4 Addresses per Interface</th>
<th>IPv6 Addresses per Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1.medium</td>
<td>2</td>
<td>6</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>c1.xlarge</td>
<td>4</td>
<td>15</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>c3.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>c3.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c3.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c3.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c3.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c4.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>c4.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c4.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c4.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c4.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c5.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>c5.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c5.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c5.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c5.9xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c5.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Instance Type</td>
<td>Maximum Network Interfaces</td>
<td>IPv4 Addresses per Interface</td>
<td>IPv6 Addresses per Interface</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>c5.18xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>c5.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>c5.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>c5d.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>c5d.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c5d.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c5d.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c5d.9xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c5d.18xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>c5n.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>c5n.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c5n.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>c5n.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c5n.9xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>c5n.18xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>c5n.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>cc2.8xlarge</td>
<td>8</td>
<td>30</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>cr1.8xlarge</td>
<td>8</td>
<td>30</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>d2.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>d2.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>d2.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>d2.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>f1.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>f1.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>f1.16xlarge</td>
<td>8</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>g2.2xlarge</td>
<td>4</td>
<td>15</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>g2.8xlarge</td>
<td>8</td>
<td>30</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>g3s.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>g3.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>g3.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Instance Type</td>
<td>Maximum Network Interfaces</td>
<td>IPv4 Addresses per Interface</td>
<td>IPv6 Addresses per Interface</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>g3.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>g4dn.xlarge</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>g4dn.2xlarge</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>g4dn.4xlarge</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>g4dn.8xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>g4dn.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>g4dn.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>h1.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>h1.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>h1.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>h1.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>hs1.8xlarge</td>
<td>8</td>
<td>30</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>i2.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>i2.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>i2.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>i2.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>i3.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>i3.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>i3.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>i3.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>i3.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>i3.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>i3.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>i3en.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>i3en.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>i3en.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>i3en.3xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>i3en.6xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>i3en.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>i3en.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Instance Type</td>
<td>Maximum Network Interfaces</td>
<td>IPv4 Addresses per Interface</td>
<td>IPv6 Addresses per Interface</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>i3en.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m1.small</td>
<td>2</td>
<td>4</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m1.medium</td>
<td>2</td>
<td>6</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m1.large</td>
<td>3</td>
<td>10</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m1.xlarge</td>
<td>4</td>
<td>15</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m2.xlarge</td>
<td>4</td>
<td>15</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m2.2xlarge</td>
<td>4</td>
<td>30</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m2.4xlarge</td>
<td>8</td>
<td>30</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m3.medium</td>
<td>2</td>
<td>6</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m3.large</td>
<td>3</td>
<td>10</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m3.xlarge</td>
<td>4</td>
<td>15</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m3.2xlarge</td>
<td>4</td>
<td>30</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>m4.large</td>
<td>2</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>m4.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m4.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m4.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m4.10xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m4.16xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>m5.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m5.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m5.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5a.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>m5a.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m5a.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Instance Type</td>
<td>Maximum Network Interfaces</td>
<td>IPv4 Addresses per Interface</td>
<td>IPv6 Addresses per Interface</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------</td>
<td>------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>m5a.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5a.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5a.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5a.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5a.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5ad.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>m5ad.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m5ad.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m5ad.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5ad.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5ad.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5d.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>m5d.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m5d.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>m5d.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5d.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5d.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>m5d.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5d.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>m5d.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>p2.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>p2.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>p2.16xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>p3.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>p3.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>p3.16xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>p3dn.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r3.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>r3.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r3.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>
## Amazon Elastic Compute Cloud
### User Guide for Windows Instances
### IP Addresses Per Network Interface Per Instance Type

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Maximum Network Interfaces</th>
<th>IPv4 Addresses per Interface</th>
<th>IPv6 Addresses per Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>r3.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r3.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r4.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>r4.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r4.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r4.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r4.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r4.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>r5.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5.large metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5a.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>r5a.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5a.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5a.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5a.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5a.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5a.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5a.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5ad.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>r5ad.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5ad.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5ad.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5ad.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>
### IP Addresses Per Network Interface Per Instance Type

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Maximum Network Interfaces</th>
<th>IPv4 Addresses per Interface</th>
<th>IPv6 Addresses per Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>r5ad.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5d.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>r5d.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5d.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>r5d.4xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5d.8xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5d.12xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>r5d.16xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5d.24xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>r5d.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>t1.micro</td>
<td>2</td>
<td>2</td>
<td>IPv6 not supported</td>
</tr>
<tr>
<td>t2.nano</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t2.micro</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t2.small</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>t2.medium</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>t2.large</td>
<td>3</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>t2.xlarge</td>
<td>3</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>t2.2xlarge</td>
<td>3</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>t3.nano</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t3.micro</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t3.small</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>t3.medium</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>t3.large</td>
<td>3</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>t3.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>t3.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>t3a.nano</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t3a.micro</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>t3a.small</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>t3a.medium</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>t3a.large</td>
<td>3</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
### Instance Type

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Maximum Network Interfaces</th>
<th>IPv4 Addresses per Interface</th>
<th>IPv6 Addresses per Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3a.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>t3a.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>u-6tb1.metal</td>
<td>5</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>u-9tb1.metal</td>
<td>5</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>u-12tb1.metal</td>
<td>5</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>x1.16xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>x1.32xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>x1e.xlarge</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>x1e.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>x1e.4xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>x1e.8xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>x1e.16xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>x1e.32xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>z1d.large</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>z1d.xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>z1d.2xlarge</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>z1d.3xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>z1d.6xlarge</td>
<td>8</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>z1d.12xlarge</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>z1d.metal</td>
<td>15</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

**Note**

If f1.16xlarge, g3.16xlarge, h1.16xlarge, i3.16xlarge, and r4.16xlarge instances use more than 31 IPv4 or IPv6 addresses per interface, they cannot access the instance metadata, VPC DNS, and Time Sync services from the 32nd IP address onwards. If access to these services is needed from all IP addresses on the interface, we recommend using a maximum of 31 IP addresses per interface.

### Scenarios for Network Interfaces

Attaching multiple network interfaces to an instance is useful when you want to:

- Create a management network.
- Use network and security appliances in your VPC.
- Create dual-homed instances with workloads/roles on distinct subnets.
- Create a low-budget, high-availability solution.
Creating a Management Network

You can create a management network using network interfaces. In this scenario, the primary network interface (eth0) on the instance handles public traffic and the secondary network interface (eth1) handles backend management traffic and is connected to a separate subnet in your VPC that has more restrictive access controls. The public interface, which may or may not be behind a load balancer, has an associated security group that allows access to the server from the internet (for example, allow TCP port 80 and 443 from 0.0.0.0/0, or from the load balancer) while the private facing interface has an associated security group allowing RDP access only from an allowed range of IP addresses either within the VPC or from the internet, a private subnet within the VPC or a virtual private gateway.

To ensure failover capabilities, consider using a secondary private IPv4 for incoming traffic on a network interface. In the event of an instance failure, you can move the interface and/or secondary private IPv4 address to a standby instance.

Use Network and Security Appliances in Your VPC

Some network and security appliances, such as load balancers, network address translation (NAT) servers, and proxy servers prefer to be configured with multiple network interfaces. You can create and attach secondary network interfaces to instances in a VPC that are running these types of applications and configure the additional interfaces with their own public and private IP addresses, security groups, and source/destination checking.

Creating Dual-homed Instances with Workloads/Roles on Distinct Subnets

You can place a network interface on each of your web servers that connects to a mid-tier network where an application server resides. The application server can also be dual-homed to a backend network (subnet) where the database server resides. Instead of routing network packets through the dual-homed instances, each dual-homed instance receives and processes requests on the front end, initiates a connection to the backend, and then sends requests to the servers on the backend network.
Create a Low Budget High Availability Solution

If one of your instances serving a particular function fails, its network interface can be attached to a replacement or hot standby instance pre-configured for the same role in order to rapidly recover the service. For example, you can use a network interface as your primary or secondary network interface to a critical service such as a database instance or a NAT instance. If the instance fails, you (or more likely, the code running on your behalf) can attach the network interface to a hot standby instance. Because the interface maintains its private IP addresses, Elastic IP addresses, and MAC address, network traffic begins flowing to the standby instance as soon as you attach the network interface to the replacement instance. Users experience a brief loss of connectivity between the time the instance fails and the time that the network interface is attached to the standby instance, but no changes to the VPC route table or your DNS server are required.

Best Practices for Configuring Network Interfaces

- You can attach a network interface to an instance when it's running (hot attach), when it's stopped (warm attach), or when the instance is being launched (cold attach).
- You can detach secondary network interfaces when the instance is running or stopped. However, you can't detach the primary network interface (eth0).
- You can move a network interface from one instance to another, if the instances are in the same Availability Zone and VPC but in different subnets.
- When launching an instance using the CLI, API, or an SDK, you can specify the primary network interface (eth0) and additional network interfaces.
- Launching an Amazon Linux or Windows Server instance with multiple network interfaces automatically configures interfaces, private IPv4 addresses, and route tables on the operating system of the instance.
- A warm or hot attach of an additional network interface may require you to manually bring up the second interface, configure the private IPv4 address, and modify the route table accordingly. Instances running Amazon Linux or Windows Server automatically recognize the warm or hot attach and configure themselves.
- Attaching another network interface to an instance (for example, a NIC teaming configuration) cannot be used as a method to increase or double the network bandwidth to or from the dual-homed instance.
- If you attach two or more network interfaces from the same subnet to an instance, you may encounter networking issues such as asymmetric routing. If possible, use a secondary private IPv4 address on the primary network interface instead. For more information, see Assigning a Secondary Private IPv4 Address (p. 664). If you need to use multiple network interfaces, you must configure the network interfaces to use static routing. For more information, see Configure a Secondary Elastic Network Interface (p. 471).

Working with Network Interfaces

You can work with network interfaces using the Amazon EC2 console or the command line.

Contents
- Creating a Network Interface (p. 691)
- Deleting a Network Interface (p. 691)
- Viewing Details about a Network Interface (p. 692)
- Attaching a Network Interface When Launching an Instance (p. 692)
- Attaching a Network Interface to a Stopped or Running Instance (p. 693)
- Detaching a Network Interface from an Instance (p. 694)
Creating a Network Interface

You can create a network interface in a subnet. You can't move the network interface to another subnet after it's created, and you can only attach the network interface to instances in the same Availability Zone.

To create a network interface using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Choose Create Network Interface.
4. For Description, enter a descriptive name.
5. For Subnet, select the subnet.
6. For Private IP (or IPv4 Private IP), enter the primary private IPv4 address. If you don't specify an IPv4 address, we select an available private IPv4 address from within the selected subnet.
7. (IPv6 only) If you selected a subnet that has an associated IPv6 CIDR block, you can optionally specify an IPv6 address in the IPv6 IP field.
8. For Security groups, select one or more security groups.
9. Choose Yes, Create.

To create a network interface using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- create-network-interface (AWS CLI)
- New-EC2NetworkInterface (AWS Tools for Windows PowerShell)

Deleting a Network Interface

To delete an instance, you must first detach the network interface. Deleting a network interface releases all attributes associated with the interface and releases any private IP addresses or Elastic IP addresses to be used by another instance.

To delete a network interface using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Select a network interface and choose Delete.
4. In the **Delete Network Interface** dialog box, choose **Yes, Delete**.

**To delete a network interface using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `delete-network-interface` (AWS CLI)
- `Remove-EC2NetworkInterface` (AWS Tools for Windows PowerShell)

**Viewing Details about a Network Interface**

You can view all the network interfaces in your account.

**To describe a network interface using the console**

1. Open the Amazon EC2 console at `https://console.aws.amazon.com/ec2/`.
2. In the navigation pane, choose **Network Interfaces**.
3. Select the network interface.
4. To view the details, choose **Details**.

**To describe a network interface using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `describe-network-interfaces` (AWS CLI)
- `Get-EC2NetworkInterface` (AWS Tools for Windows PowerShell)

**To describe a network interface attribute using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `describe-network-interface-attribute` (AWS CLI)

**Attaching a Network Interface When Launching an Instance**

You can specify an existing network interface or attach an additional network interface when you launch an instance.

**Note**

If an error occurs when attaching a network interface to your instance, this causes the instance launch to fail.

**To attach a network interface when launching an instance using the console**

1. Open the Amazon EC2 console at `https://console.aws.amazon.com/ec2/`.
2. Choose **Launch Instance**.
3. Select an AMI and instance type and choose **Next: Configure Instance Details**.
4. On the **Configure Instance Details** page, select a VPC for **Network**, and a subnet for **Subnet**.
5. In the **Network Interfaces** section, the console enables you to specify up to two network interfaces (new, existing, or a combination) when you launch an instance. You can also enter a primary IPv4 address and one or more secondary IPv4 addresses for any new interface.

You can add additional network interfaces to the instance after you launch it. The total number of network interfaces that you can attach varies by instance type. For more information, see **IP Addresses Per Network Interface Per Instance Type** (p. 681).

**Note**
If you specify more than one network interface, you cannot auto-assign a public IPv4 address to your instance.

6. (IPv6 only) If you’re launching an instance into a subnet that has an associated IPv6 CIDR block, you can specify IPv6 addresses for any network interfaces that you attach. Under **IPv6 IPs**, choose **Add IP**. To add a secondary IPv6 address, choose **Add IP** again. You can enter an IPv6 address from the range of the subnet, or leave the default **Auto-assign** value to let Amazon choose an IPv6 address from the subnet for you.

7. Choose **Next: Add Storage**.

8. On the **Add Storage** page, you can specify volumes to attach to the instance besides the volumes specified by the AMI (such as the root device volume), and then choose **Next: Add Tags**.

9. On the **Add Tags** page, specify tags for the instance, such as a user-friendly name, and then choose **Next: Configure Security Group**.

10. On the **Configure Security Group** page, you can select a security group or create a new one. Choose **Review and Launch**.

**Note**
If you specified an existing network interface in step 5, the instance is associated with the security group for that network interface, regardless of any option that you select in this step.

11. On the **Review Instance Launch** page, details about the primary and additional network interface are displayed. Review the settings, and then choose **Launch** to choose a key pair and launch your instance. If you’re new to Amazon EC2 and haven’t created any key pairs, the wizard prompts you to create one.

**To attach a network interface when launching an instance using the command line**

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).

- `run-instances` (AWS CLI)
- `New-EC2Instance` (AWS Tools for Windows PowerShell)

**Attaching a Network Interface to a Stopped or Running Instance**

You can attach a network interface to any of your stopped or running instances in your VPC, using either the **Instances** or **Network Interfaces** pages of the Amazon EC2 console.

**Note**
If the public IPv4 address on your instance is released, it does not receive a new one if there is more than one network interface attached to the instance. For more information about the behavior of public IPv4 addresses, see **Public IPv4 Addresses and External DNS Hostnames** (p. 657).

**To attach a network interface to an instance using the Instances page**

1. Open the Amazon EC2 console at `https://console.aws.amazon.com/ec2/`. 
2. In the navigation pane, choose **Instances**.
3. Choose **Actions, Networking, Attach Network Interface**.
4. In the **Attach Network Interface** dialog box, select the network interface and choose **Attach**.

### To attach a network interface to an instance using the Network Interfaces page

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Network Interfaces**.
3. Select the network interface and choose **Attach**.
4. In the **Attach Network Interface** dialog box, select the instance and choose **Attach**.

### To attach a network interface to an instance using the command line

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).

- `attach-network-interface` (AWS CLI)
- `Add-EC2NetworkInterface` (AWS Tools for Windows PowerShell)

### Detaching a Network Interface from an Instance

You can detach a secondary network interface at any time, using either the **Instances** or **Network Interfaces** page of the Amazon EC2 console.

#### To detach a network interface from an instance using the Instances page

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Instances**.
3. Choose **Actions, Networking, Detach Network Interface**.
4. In the **Detach Network Interface** dialog box, select the network interface and choose **Detach**.

#### To detach a network interface from an instance using the Network Interfaces page

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Network Interfaces**.
3. Select the network interface and choose **Detach**.
4. In the **Detach Network Interface** dialog box, choose **Yes, Detach**. If the network interface fails to detach from the instance, choose **Force detachment**, and then try again.

**Note**

- Use the **Force detachment** option only as a last resort to detach a network interface from a failed instance.
- If you use the **Force detachment** option to detach a network interface, you might not be able to attach a different network interface to the same index on the instance without first stopping and starting the instance.
- If you force the detachment of a network interface, the **instance metadata** (p. 478) might not get updated. This means that the attributes associated with the detached network interface might still be visible. The instance metadata will get updated when you stop and start the instance.
- You can't force the detachment of a network interface from EC2-Classic instances.
To detach a network interface using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- detach-network-interface (AWS CLI)
- Dismount-EC2NetworkInterface (AWS Tools for Windows PowerShell)

Changing the Security Group

You can change the security groups that are associated with a network interface. When you create the security group, be sure to specify the same VPC as the subnet for the network interface.

Note
To change security group membership for interfaces owned by other services, such as Elastic Load Balancing, use the console or command line interface for that service.

To change the security group of a network interface using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
4. In the Change Security Groups dialog box, select the security groups to use, and choose Save.

To change the security group of a network interface using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- modify-network-interface-attribute (AWS CLI)
- Edit-EC2NetworkInterfaceAttribute (AWS Tools for Windows PowerShell)

Changing the Source or Destination Checking

The Source/Destination Check attribute controls whether source/destination checking is enabled on the instance. Disabling this attribute enables an instance to handle network traffic that isn't specifically destined for the instance. For example, instances running services such as network address translation, routing, or a firewall should set this value to disabled. The default value is enabled.

To change source/destination checking for a network interface using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Select the network interface and choose Actions, Change Source/Dest Check.
4. In the dialog box, choose Enabled (if enabling) or Disabled (if disabling), and Save.

To change source/destination checking for a network interface using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- modify-network-interface-attribute (AWS CLI)
• Edit-EC2NetworkInterfaceAttribute (AWS Tools for Windows PowerShell)

**Associating an Elastic IP Address (IPv4)**

If you have an Elastic IP address (IPv4), you can associate it with one of the private IPv4 addresses for the network interface. You can associate one Elastic IP address with each private IPv4 address.

You can associate an Elastic IP address using the Amazon EC2 console or the command line.

**To associate an Elastic IP address using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Network Interfaces**.
3. Select the network interface and choose **Actions**, **Associate Address**.
4. In the **Associate Elastic IP Address** dialog box, select the Elastic IP address from the **Address** list.
5. For **Associate to private IP address**, select the private IPv4 address to associate with the Elastic IP address.
6. Choose **Allow reassociation** to allow the Elastic IP address to be associated with the specified network interface if it's currently associated with another instance or network interface, and then choose **Associate Address**.

**To associate an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

• associate-address (AWS CLI)
• Register-EC2Address (AWS Tools for Windows PowerShell)

**Disassociating an Elastic IP Address (IPv4)**

If the network interface has an Elastic IP address (IPv4) associated with it, you can disassociate the address, and then either associate it with another network interface or release it back to the address pool. This is the only way to associate an Elastic IP address with an instance in a different subnet or VPC using a network interface, as network interfaces are specific to a particular subnet.

You can disassociate an Elastic IP address using the Amazon EC2 console or the command line.

**To disassociate an Elastic IP address using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Network Interfaces**.
3. Select the network interface and choose **Actions**, **Disassociate Address**.
4. In the **Disassociate IP Address** dialog box, choose **Yes, Disassociate**.

**To disassociate an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

• disassociate-address (AWS CLI)
• Unregister-EC2Address (AWS Tools for Windows PowerShell)
Assigning an IPv6 Address

You can assign one or more IPv6 addresses to a network interface. The network interface must be in a subnet that has an associated IPv6 CIDR block. To assign a specific IPv6 address to the network interface, ensure that the IPv6 address is not already assigned to another network interface.

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces and select the network interface.
3. Choose Actions, Manage IP Addresses.
4. Under IPv6 Addresses, choose Assign new IP. Specify an IPv6 address from the range of the subnet. To let AWS choose an address for you, leave the Auto-assign value.
5. Choose Yes, Update.

To assign an IPv6 address to a network interface using the command line

- You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
  - assign-ipv6-addresses (AWS CLI)
  - Register-EC2ipv6AddressList (AWS Tools for Windows PowerShell)

Unassigning an IPv6 Address

You can unassign an IPv6 address from a network interface using the Amazon EC2 console.

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces and select the network interface.
3. Choose Actions, Manage IP Addresses.
4. Under IPv6 Addresses, choose Unassign for the IPv6 address to remove.
5. Choose Yes, Update.

To unassign an IPv6 address from a network interface using the command line

- You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
  - unassign-ipv6-addresses (AWS CLI)
  - Unregister-EC2ipv6AddressList (AWS Tools for Windows PowerShell)

Changing Termination Behavior

You can set the termination behavior for a network interface that's attached to an instance. You can specify whether the network interface should be automatically deleted when you terminate the instance to which it's attached.

You can change the terminating behavior for a network interface using the Amazon EC2 console or the command line.

To change the termination behavior for a network interface using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Select the network interface and choose Actions, Change Termination Behavior.
4. In the Change Termination Behavior dialog box, select the Delete on termination check box if you want the network interface to be deleted when you terminate an instance.

To change the termination behavior for a network interface using the command line
You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- modify-network-interface-attribute (AWS CLI)
- Edit-EC2NetworkInterfaceAttribute (AWS Tools for Windows PowerShell)

Adding or Editing a Description
You can change the description for a network interface using the Amazon EC2 console or the command line.

To change the description for a network interface using the console
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Select the network interface and choose Actions, Change Description.
4. In the Change Description dialog box, enter a description for the network interface, and then choose Save.

To change the description for a network interface using the command line
You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- modify-network-interface-attribute (AWS CLI)
- Edit-EC2NetworkInterfaceAttribute (AWS Tools for Windows PowerShell)

Adding or Editing Tags
Tags are metadata that you can add to a network interface. Tags are private and are only visible to your account. Each tag consists of a key and an optional value. For more information about tags, see Tagging Your Amazon EC2 Resources (p. 964).

To add or edit tags for a network interface using the console
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Select the network interface.
4. In the details pane, choose Tags, Add/Edit Tags.
5. In the Add/Edit Tags dialog box, choose Create Tag for each tag to create, and enter a key and optional value. When you're done, choose Save.

To add or edit tags for a network interface using the command line
You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
Requester-Managed Network Interfaces

A requester-managed network interface is a network interface that an AWS service creates in your VPC. This network interface can represent an instance for another service, such as an Amazon RDS instance, or it can enable you to access another service or resource, such as an AWS PrivateLink service, or an Amazon ECS task.

You cannot modify or detach a requester-managed network interface. If you delete the resource that the network interface represents, the AWS service detaches and deletes the network interface for you. To change the security groups for a requester-managed network interface, you might have to use the console or command line tools for that service. For more information, see the service-specific documentation.

You can tag a requester-managed network interface. For more information, see Adding or Editing Tags (p. 698).

You can view the requester-managed network interfaces that are in your account.

To view requester-managed network interfaces using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Network Interfaces.
3. Select the network interface and view the following information on the details pane:
   - **Attachment owner**: If you created the network interface, this field displays your AWS account ID. Otherwise, it displays an alias or ID for the principal or service that created the network interface.
   - **Description**: Provides information about the purpose of the network interface; for example, "VPC Endpoint Interface".

To view requester-managed network interfaces using the command line

1. Use the describe-network-interfaces AWS CLI command to describe the network interfaces in your account.

   ```bash
   aws ec2 describe-network-interfaces
   ```

2. In the output, the RequesterManaged field displays true if the network interface is managed by another AWS service.

   ```json
   {
     "Status": "in-use",
     ...
     "Description": "VPC Endpoint Interface vpce-089f2123488812123",
     "NetworkInterfaceId": "eni-c8fbc27e",
     "VpcId": "vpc-1a2b3c4d",
     "PrivateIpAddresses": [
       {
         "PrivateDnsName": "ip-10-0-2-227.ec2.internal",
         "Primary": true,
         "PrivateIpAddress": "10.0.2.227"
       }
     ],
     "RequesterManaged": true,
     ...
   }
   ```

## Enhanced Networking on Windows

Enhanced networking uses single root I/O virtualization (SR-IOV) to provide high-performance networking capabilities on supported instance types (p. 700). SR-IOV is a method of device virtualization that provides higher I/O performance and lower CPU utilization when compared to traditional virtualized network interfaces. Enhanced networking provides higher bandwidth, higher packet per second (PPS) performance, and consistently lower inter-instance latencies. There is no additional charge for using enhanced networking.

### Contents
- Enhanced Networking Types (p. 700)
- Enabling Enhanced Networking on Your Instance (p. 700)
- Enabling Enhanced Networking with the Elastic Network Adapter (ENA) on Windows Instances (p. 701)
- Enabling Enhanced Networking with the Intel 82599 VF Interface on Windows Instances (p. 707)

### Enhanced Networking Types

Depending on your instance type, enhanced networking can be enabled using one of the following mechanisms:

**Elastic Network Adapter (ENA)**

The Elastic Network Adapter (ENA) supports network speeds of up to 100 Gbps for supported instance types.

C5, C5d, C5n, F1, G3, G4, H1, I3, I3en, m4.16xlarge, M5, M5a, M5ad, M5d P2, P3, R4, R5, R5a, R5ad, R5d, T3, T3a, u-6tb1.metal, u-9tb1.metal, u-12tb1.metal, X1, X1e, and z1d instances use the Elastic Network Adapter for enhanced networking.

**Intel 82599 Virtual Function (VF) interface**

The Intel 82599 Virtual Function interface supports network speeds of up to 10 Gbps for supported instance types.

C3, C4, D2, I2, M4 (excluding m4.16xlarge), and R3 instances use the Intel 82599 VF interface for enhanced networking.

For information about the supported network speed for each instance type, see Amazon EC2 Instance Types.

### Enabling Enhanced Networking on Your Instance

If your instance type supports the Elastic Network Adapter for enhanced networking, follow the procedures in Enabling Enhanced Networking with the Elastic Network Adapter (ENA) on Windows Instances (p. 701).

If your instance type supports the Intel 82599 VF interface for enhanced networking, follow the procedures in Enabling Enhanced Networking with the Intel 82599 VF Interface on Windows Instances (p. 707).
Enabling Enhanced Networking with the Elastic Network Adapter (ENA) on Windows Instances

Amazon EC2 provides enhanced networking capabilities through the Elastic Network Adapter (ENA).

Contents
- Requirements (p. 701)
- Data Encryption (p. 701)
- Testing Whether Enhanced Networking Is Enabled (p. 701)
- Enabling Enhanced Networking on Windows (p. 702)
- Amazon ENA Driver Versions (p. 703)
- Subscribing to Notifications (p. 435)
- Operating System Optimizations (p. 706)

Requirements

To prepare for enhanced networking using the ENA, set up your instance as follows:

- Select from the following supported instance types: C5, C5d, C5n, F1, G3, G4, H1, I3, I3en, m4.16xlarge, M5, M5a, M5ad, M5d P2, P3, R4, R5, R5a, R5ad, R5d, T3, T3a, u-6tb1.metal, u-9tb1.metal, u-12tb1.metal, X1, X1e, and z1d.
- Ensure that the instance has internet connectivity.
- Install and configure the AWS CLI or the AWS Tools for Windows PowerShell on any computer you choose, preferably your local desktop or laptop. For more information, see Accessing Amazon EC2 (p. 3). Enhanced networking cannot be managed from the Amazon EC2 console.
- If you have important data on the instance that you want to preserve, you should back that data up now by creating an AMI from your instance. Updating kernels and kernel modules, as well as enabling the enaSupport attribute, might render incompatible instances or operating systems unreachable; if you have a recent backup, your data will still be retained if this happens.

Data Encryption

AWS provides secure and private connectivity between EC2 instances. In addition, we automatically encrypt in-transit traffic between supported instances in the same VPC or in peered VPCs, using AEAD algorithms with 256-bit encryption. This encryption feature uses the offload capabilities of the underlying hardware, and there is no impact on network performance. The supported instances are: C5n, G4, I3en, P3dn.

Testing Whether Enhanced Networking Is Enabled

To test whether enhanced networking is already enabled, verify that the driver is installed on your instance and that the enaSupport attribute is set.

Instance Attribute (enaSupport)

To check whether an instance has the enhanced networking enaSupport attribute set, use one of the following commands. If the attribute is set, the response is true.

- describe-instances (AWS CLI)
Image Attribute (enaSupport)

To check whether an AMI has the enhanced networking enaSupport attribute set, use one of the following commands. If the attribute is set, the response is true.

- **describe-images** (AWS CLI)

```bash
aws ec2 describe-images --image-id ami_id --query "Images[].EnaSupport"
```

- **Get-EC2Image** (Tools for Windows PowerShell)

```powershell
(Get-EC2Image -ImageId ami_id).EnaSupport
```

Enabling Enhanced Networking on Windows

If you launched your instance and it does not have enhanced networking enabled already, you must download and install the required network adapter driver on your instance, and then set the enaSupport instance attribute to activate enhanced networking. You can only enable this attribute on supported instance types and only if the ENA driver is installed. For more information, see Enhanced Networking Types (p. 700).

**To enable enhanced networking**

1. Connect to your instance and log in as the local administrator.
2. [Windows Server 2016 and later] Run the following EC2Launch PowerShell script to configure the instance after the driver is installed.

   ```powershell
   C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeInstance.ps1 - Schedule
   ```

3. From the instance, install the driver as follows:
   a. Download the latest driver to the instance.
   b. Extract the zip archive.
   c. Install the driver by running the `install.ps1` PowerShell script.
4. From your local computer, stop the instance using the Amazon EC2 console or one of the following commands: `stop-instances` (AWS CLI), `Stop-EC2Instance` (AWS Tools for Windows PowerShell). If your instance is managed by AWS OpsWorks, you should stop the instance in the AWS OpsWorks console so that the instance state remains in sync.
5. Enable ENA support on your instance as follows:
   a. From your local computer, check the EC2 instance ENA support attribute on your instance by running one of the following commands. If the attribute is not enabled, the output will be "[]" or blank. `EnaSupport` is set to `false` by default.

   ```bash
   aws ec2 describe-instances --instance-ids instance_id --query "Reservations[].Instances[].EnaSupport"
   ```

   ```powershell
   (Get-EC2Instance -InstanceId instance-id).Instances.EnaSupport
   ```

   • **describe-instances** (AWS CLI)
Enhanced Networking: ENA

- AWS EC2 CLI
  ```bash
  aws ec2 describe-instances --instance-ids instance_id --query "Reservations[0].Instances[0].EnaSupport"
  ```
- Get-EC2Instance (Tools for Windows PowerShell)
  ```powershell
  (Get-EC2Instance -InstanceId instance-id).Instances.EnaSupport
  ```

b. To enable ENA support, run one of the following commands:

- modify-instance-attribute (AWS CLI)
  ```bash
  aws ec2 modify-instance-attribute --instance-id instance_id --ena-support
  ```
- Edit-EC2InstanceAttribute (AWS Tools for Windows PowerShell)
  ```powershell
  Edit-EC2InstanceAttribute -InstanceId instance_id -EnaSupport $true
  ```

If you encounter problems when you restart the instance, you can also disable ENA support using one of the following commands:

- modify-instance-attribute (AWS CLI)
  ```bash
  aws ec2 modify-instance-attribute --instance-id instance_id --no-ena-support
  ```
- Edit-EC2InstanceAttribute (AWS Tools for Windows PowerShell)
  ```powershell
  Edit-EC2InstanceAttribute -InstanceId instance_id -EnaSupport $false
  ```
c. Verify that the attribute has been set to true using describe-instances or Get-EC2Instance as shown previously. You should now see the following output:

```plaintext
[ true ]
```

6. From your local computer, start the instance using the Amazon EC2 console or one of the following commands: start-instances (AWS CLI), Start-EC2Instance (AWS Tools for Windows PowerShell). If your instance is managed by AWS OpsWorks, you should start the instance using the AWS OpsWorks console so that the instance state remains in sync.

7. On the instance, validate that the ENA driver is installed and enabled as follows:

a. Right-click the network icon and choose Open Network and Sharing Center.

b. Choose the Ethernet adapter (for example, Ethernet 2).

c. Choose Details. For Network Connection Details, check that Description is Amazon Elastic Network Adapter.

8. (Optional) Create an AMI from the instance. The AMI inherits the enaSupport attribute from the instance. Therefore, you can use this AMI to launch another instance with ENA enabled by default. For more information, see Creating a Custom Windows AMI (p. 64).

### Amazon ENA Driver Versions

Windows AMIs include the Amazon ENA driver to enable enhanced networking. The following table summarizes the changes for each release.
<table>
<thead>
<tr>
<th>Driver version</th>
<th>Details</th>
<th>Release date</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1</td>
<td>Bug Fixes</td>
<td>September 2019</td>
</tr>
<tr>
<td></td>
<td>• Prevent drops of highly fragmented TCP LSO packets arriving from operating system.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Properly handle Encapsulating Security Payload (ESP) protocol within the IPSec in IPv6 networks.</td>
<td></td>
</tr>
<tr>
<td>2.1.0</td>
<td>ENA Windows driver v2.1 introduces new ENA device capabilities, provides a performance boost, adds new features, and includes multiple stability improvements.</td>
<td>July 2019</td>
</tr>
<tr>
<td></td>
<td>• New features</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Use standardized Windows registry key for Jumbo frames configuration.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Allow VLAN ID setting via the ENA driver properties GUI.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Improved Recovery flows</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Improved failure identification mechanism.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Added support for tunable recovery parameters.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Support up to 32 I/O queues for newer EC2 instances that have more than 8 vCPUs.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• ~90% reduction of driver memory footprint.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Performance optimizations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Reduced transmit path latency.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Support for receive checksum offload.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Performance optimization for heavily loaded system (optimized usage of locking mechanisms).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Further enhancements to reduce CPU utilization and improve system responsiveness under load.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Bug Fixes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fix crash due to invalid parsing of non-contiguous Tx headers.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fix driver v1.5 crash during ENI detach on Bare Metal instances.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fix LSO pseudo-header checksum calculation error over IPv6.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fix potential memory resource leak upon initialization failure.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Disable TCP/UDP checksum offload for IPv4 fragments.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fix for VLAN configuration. VLAN was incorrectly disabled when only VLAN priority should have been disabled.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Enable correct parsing of custom driver messages by the event viewer.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fix failure to initialize driver due to invalid timestamp handling.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Fix race condition between data processing and ENA device disabling.</td>
<td></td>
</tr>
<tr>
<td>1.5.0</td>
<td>• Improved stability and performance fixes.</td>
<td>October 2018</td>
</tr>
</tbody>
</table>
Enhanced Networking: ENA

<table>
<thead>
<tr>
<th>Driver version</th>
<th>Details</th>
<th>Release date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2.3</td>
<td>• Receive Buffers can now be configured up to a value of 8192 in Advanced Properties of the ENA NIC. • Default Receive Buffers of 1k. Includes reliability fixes and unifies support for Windows Server 2008 R2 through Windows Server 2016.</td>
<td>February 2018</td>
</tr>
<tr>
<td>1.0.9</td>
<td>Includes some reliability fixes. Applies only to Windows Server 2008 R2. Not recommended for other versions of Windows Server.</td>
<td>December 2016</td>
</tr>
</tbody>
</table>

Subscribing to Notifications

Amazon SNS can notify you when new versions of EC2 Windows Drivers are released. Use the following procedure to subscribe to these notifications.

To subscribe to EC2 notifications

2. In the navigation bar, change the region to US East (N. Virginia), if necessary. You must select this region because the SNS notifications that you are subscribing to are in this region.
3. In the navigation pane, choose Subscriptions.
4. Choose Create subscription.
5. In the Create subscription dialog box, do the following:
   a. For TopicARN, copy the following Amazon Resource Name (ARN):
      arn:aws:sns:us-east-1:801119661308:ec2-windows-drivers
   b. For Protocol, choose Email.
   c. For Endpoint, type an email address that you can use to receive the notifications.
   d. Choose Create subscription.
6. You’ll receive a confirmation email. Open the email and follow the directions to complete your subscription.

Whenever new EC2 Windows drivers are released, we send notifications to subscribers. If you no longer want to receive these notifications, use the following procedure to unsubscribe.

To unsubscribe from Amazon EC2 Windows driver notification

2. In the navigation pane, choose Subscriptions.
3. Select the checkbox for the subscription and then choose Actions, Delete subscriptions. When prompted for confirmation, choose Delete.
Operating System Optimizations

To achieve the maximum network performance on instances with enhanced networking, you may need to modify the default operating system configuration. We recommend the following configuration changes for applications that require high network performance. Other optimizations (such as turning on checksum offloading and enabling RSS, for example) are already in place on official Windows AMIs.

**Note**
TCP chimney offloading should be disabled in most use cases, and has been deprecated as of Windows Server 2016.

In addition to these operating system optimizations, you should also consider the maximum transmission unit (MTU) of your network traffic, and adjust according to your workload and network architecture. For more information, see Network Maximum Transmission Unit (MTU) for Your EC2 Instance (p. 718).

AWS regularly measures average round trip latencies between instances launched in a cluster placement group of 50us and tail latencies of 200us at the 99.9 percentile. If your applications require consistently low latencies, we recommend using the latest version of the ENA drivers on fixed performance Nitro-based instances.

Configure RSS CPU Affinity

Receive side scaling (RSS) is used to distribute network traffic CPU load across multiple processors. By default, the official Amazon Windows AMIs are configured with RSS enabled. ENA ENIs provide up to eight RSS queues. By defining CPU affinity for RSS queues, as well as for other system processes, it is possible to spread the CPU load out over multi-core systems, enabling more network traffic to be processed. On instance types with more than 16 vCPUs, we recommend you use the `Set-NetAdapterRSS` PowerShell cmdlet (available from Windows Server 2012 and later), which manually excludes the boot processor (logical processor 0 and 1 when hyper-threading is enabled) from the RSS configuration for all ENIs, in order to prevent contention with various system components.

Windows is hyper-thread aware and will ensure the RSS queues of a single NIC are always placed on different physical cores. Therefore, unless hyper-threading is disabled, in order to completely prevent contention with other NICs, spread the RSS configuration of each NIC between a range of 16 logical processors. The `Set-NetAdapterRSS` cmdlet allows you to define the per-NIC range of valid logical processors by defining the values of BaseProcessorGroup, BaseProcessorNumber, MaxProcessingGroup, MaxProcessorNumber, and NumaNode (optional). If there are not enough physical cores to completely eliminate inter-NIC contention, minimize the overlapping ranges or reduce the number of logical processors in the ENI ranges depending on the expected workload of the ENI (in other words, a low volume admin network ENI may not need as many RSS queues assigned). Also, as noted above, various components must run on CPU 0, and therefore we recommend excluding it from all RSS configurations when sufficient vCPUs are available.

For example, when there are three ENIs on a 72 vCPU instance with 2 NUMA nodes with hyper-threading enabled, the following commands spread the network load between the two CPUs without overlap and prevent the use of core 0 completely.

```
Set-NetAdapterRSS -Name NIC1 -BaseProcessorGroup 0 -BaseProcessorNumber 2 -MaxProcessorNumber 16
Set-NetAdapterRSS -Name NIC2 -BaseProcessorGroup 1 -BaseProcessorNumber 0 -MaxProcessorNumber 14
Set-NetAdapterRSS -Name NIC3 -BaseProcessorGroup 1 -BaseProcessorNumber 16 -MaxProcessorNumber 30
```

Note that these settings are persistent for each network adapter. If an instance is resized to one with a different number of vCPUs, you should reevaluate the RSS configuration for each enabled ENI. The complete Microsoft documentation for the `Set-NetAdapterRSS` cmdlet can be found here: [https://docs.microsoft.com/en-us/powershell/module/netadapter/set-netadapterrss](https://docs.microsoft.com/en-us/powershell/module/netadapter/set-netadapterrss).
Special note for SQL workloads: We also recommend that you review your IO thread affinity settings along with your ENI RSS configuration to minimize IO and network contention for the same CPUs. See affinity mask Server Configuration Option.

Enabling Enhanced Networking with the Intel 82599 VF Interface on Windows Instances

Amazon EC2 provides enhanced networking capabilities through the Intel 82599 VF interface, which uses the Intel `ixgbevf` driver.

Contents

• Requirements (p. 707)
• Testing Whether Enhanced Networking is Enabled (p. 707)
• Enabling Enhanced Networking on Windows (p. 708)

Requirements

To prepare for enhanced networking using the Intel 82599 VF interface, set up your instance as follows:

• Select from the following supported instance types: C3, C4, D2, I2, M4 (excluding m4.16xlarge), and R3.
• Launch the instance from a 64-bit HVM AMI. You can't enable enhanced networking on Windows Server 2008 and Windows Server 2003. Enhanced networking is already enabled for Windows Server 2012 R2 and Windows Server 2016 and later AMIs. Windows Server 2012 R2 includes Intel driver 1.0.15.3 and we recommend that you upgrade that driver to the latest version using the Pnputil.exe utility.
• Ensure that the instance has internet connectivity.
• Install and configure the AWS CLI or the AWS Tools for Windows PowerShell on any computer you choose, preferably your local desktop or laptop. For more information, see Accessing Amazon EC2 (p. 3). Enhanced networking cannot be managed from the Amazon EC2 console.
• If you have important data on the instance that you want to preserve, you should back that data up now by creating an AMI from your instance. Updating kernels and kernel modules, as well as enabling the `sriovNetSupport` attribute, might render incompatible instances or operating systems unreachable; if you have a recent backup, your data will still be retained if this happens.

Testing Whether Enhanced Networking is Enabled

Enhanced networking with the Intel 82599 VF interface is enabled if the driver is installed on your instance and the `sriovNetSupport` attribute is set.

Driver

To verify that the driver is installed, connect to your instance and open Device Manager. You should see "Intel(R) 82599 Virtual Function" listed under Network adapters.

Instance Attribute (sriovNetSupport)

To check whether an instance has the enhanced networking `sriovNetSupport` attribute set, use one of the following commands:

• `describe-instance-attribute` (AWS CLI)

```bash
aws ec2 describe-instance-attribute --instance-id instance_id --attribute sriovNetSupport
```
• **Get-EC2InstanceAttribute** (AWS Tools for Windows PowerShell)

```powershell
Get-EC2InstanceAttribute -InstanceId instance-id -Attribute sriovNetSupport
```

If the attribute isn't set, `SriovNetSupport` is empty; otherwise, it is set as follows:

```json
"SriovNetSupport": {
  "Value": "simple"
},
```

### Image Attribute (sriovNetSupport)

To check whether an AMI already has the enhanced networking `sriovNetSupport` attribute set, use one of the following commands:

- **describe-image-attribute** (AWS CLI)

```bash
aws ec2 describe-image-attribute --image-id ami_id --attribute sriovNetSupport
```

Note that this command only works for images that you own. You receive an `AuthFailure` error for images that do not belong to your account.

- **Get-EC2ImageAttribute** (AWS Tools for Windows PowerShell)

```powershell
Get-EC2ImageAttribute -ImageId ami-id -Attribute sriovNetSupport
```

If the attribute isn't set, `SriovNetSupport` is empty; otherwise, it is set as follows:

```json
"SriovNetSupport": {
  "Value": "simple"
},
```

### Enabling Enhanced Networking on Windows

If you launched your instance and it does not have enhanced networking enabled already, you must download and install the required network adapter driver on your instance, and then set the `sriovNetSupport` instance attribute to activate enhanced networking. You can only enable this attribute on supported instance types. For more information, see Enhanced Networking Types (p. 700).

**Important**

To view the latest version of the Intel driver in the Windows AMIs, see Details About AWS Windows AMI Versions (p. 82).

**Warning**

There is no way to disable the enhanced networking attribute after you've enabled it.

### To enable enhanced networking

1. Connect to your instance and log in as the local administrator.
2. [Windows Server 2016 and later] Run the following EC2Launch PowerShell script to configure the instance after the driver is installed.

```powershell
PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeInstance.ps1 -Schedule
```
Important
The administrator password will reset when you enable the initialize instance EC2Launch script. You can modify the configuration file to disable the administrator password reset by specifying it in the settings for the initialization tasks. For steps on how to disable password reset, see Configure Initialization Tasks (p. 397).

3. From the instance, install the driver as follows:
   a. Download the Intel network adapter driver for your operating system:
      - Windows Server 2008 R2
      - Windows Server 2012
      - Windows Server 2012 R2
      - Windows Server 2016 (including for Server version 1803 and earlier*)
      - Windows Server 2019 (including for Server version 1809 and later*)

      *Server versions 1803 and earlier as well as 1809 and later are not specifically addressed on the Intel Drivers and Software pages.

   b. In the Download folder, locate the PROWinx64.exe file. Rename this file PROWinx64.zip.
   c. Open a context (right-click) menu on PROWinx64.zip and choose Extract All. Specify a destination path and choose Extract.
   d. Open a command prompt window, go to the folder with the extracted files, and use the pnputil utility to add and install the INF file in the driver store.

      Windows Server 2019
      ```
      pnputil -i -a PROXGB\Winx64\NDIS68\vxn68x64.inf
      ```

      Windows Server 2016
      ```
      pnputil -i -a PROXGB\Winx64\NDIS65\vxn65x64.inf
      ```

      Windows Server 2012 R2
      ```
      pnputil -i -a PROXGB\Winx64\NDIS64\vxn64x64.inf
      ```

      Windows Server 2012
      ```
      pnputil -i -a PROXGB\Winx64\NDIS63\vxn63x64.inf
      ```

      Windows Server 2008 R2
      ```
      pnputil -a PROXGB\Winx64\NDIS62\vxn62x64.inf
      ```

4. From your local computer, stop the instance using the Amazon EC2 console or one of the following commands: stop-instances (AWS CLI), Stop-EC2Instance (AWS Tools for Windows PowerShell). If your instance is managed by AWS OpsWorks, you should stop the instance in the AWS OpsWorks console so that the instance state remains in sync.

5. From your local computer, enable the enhanced networking attribute using one of the following commands:
   - modify-instance-attribute (AWS CLI)
Placement Groups

When you launch a new EC2 instance, the EC2 service attempts to place the instance in such a way that all of your instances are spread out across underlying hardware to minimize correlated failures. You can use placement groups to influence the placement of a group of interdependent instances to meet the needs of your workload. Depending on the type of workload, you can create a placement group using one of the following placement strategies:

- **Cluster** – packs instances close together inside an Availability Zone. This strategy enables workloads to achieve the low-latency network performance necessary for tightly-coupled node-to-node communication that is typical of HPC applications.
- **Partition** – spreads your instances across logical partitions such that groups of instances in one partition do not share the underlying hardware with groups of instances in different partitions. This strategy is typically used by large distributed and replicated workloads, such as Hadoop, Cassandra, and Kafka.
- **Spread** – strictly places a small group of instances across distinct underlying hardware to reduce correlated failures.

There is no charge for creating a placement group.

Contents

- Cluster Placement Groups (p. 710)
- Partition Placement Groups (p. 711)
- Spread Placement Groups (p. 712)
- Placement Group Rules and Limitations (p. 713)
- Creating a Placement Group (p. 714)
- Launching Instances in a Placement Group (p. 715)
- Describing Instances in a Placement Group (p. 716)
- Changing the Placement Group for an Instance (p. 717)
- Deleting a Placement Group (p. 718)

Cluster Placement Groups

A cluster placement group is a logical grouping of instances within a single Availability Zone. A placement group can span peered VPCs in the same Region. The chief benefit of a cluster placement...
group, in addition to a 10 Gbps flow limit, is the non-blocking, non-oversubscribed, fully bi-sectional nature of the connectivity. In other words, all nodes within the placement group can talk to all other nodes within the placement group at the full line rate of 10 Gbps flows and 100 Gbps aggregate without any slowing due to over-subscription.

The following image shows instances that are placed into a cluster placement group.

Cluster placement groups are recommended for applications that benefit from low network latency, high network throughput, or both, and if the majority of the network traffic is between the instances in the group. To provide the lowest latency and the highest packet-per-second network performance for your placement group, choose an instance type that supports enhanced networking. For more information, see Enhanced Networking (p. 700).

We recommend that you launch the number of instances that you need in the placement group in a single launch request and that you use the same instance type for all instances in the placement group. If you try to add more instances to the placement group later, or if you try to launch more than one instance type in the placement group, you increase your chances of getting an insufficient capacity error.

If you stop an instance in a placement group and then start it again, it still runs in the placement group. However, the start fails if there isn’t enough capacity for the instance.

If you receive a capacity error when launching an instance in a placement group that already has running instances, stop and start all of the instances in the placement group, and try the launch again. Restarting the instances may migrate them to hardware that has capacity for all the requested instances.

**Partition Placement Groups**

Partition placement groups help reduce the likelihood of correlated hardware failures for your application. When using partition placement groups, Amazon EC2 divides each group into logical segments called partitions. Amazon EC2 ensures that each partition within a placement group has its own set of racks. Each rack has its own network and power source. No two partitions within a placement group share the same racks, allowing you to isolate the impact of hardware failure within your application.

The following image is a simple visual representation of a partition placement group in a single Availability Zone. It shows instances that are placed into a partition placement group with three partitions—**Partition 1**, **Partition 2**, and **Partition 3**. Each partition comprises multiple instances. The instances in a partition do not share racks with the instances in the other partitions, allowing you to contain the impact of a single hardware failure to only the associated partition.
Partition placement groups can be used to deploy large distributed and replicated workloads, such as HDFS, HBase, and Cassandra, across distinct racks. When you launch instances into a partition placement group, Amazon EC2 tries to distribute the instances evenly across the number of partitions that you specify. You can also launch instances into a specific partition to have more control over where the instances are placed.

A partition placement group can have partitions in multiple Availability Zones in the same Region. A partition placement group can have a maximum of seven partitions per Availability Zone. The number of instances that can be launched into a partition placement group is limited only by the limits of your account.

In addition, partition placement groups offer visibility into the partitions—you can see which instances are in which partitions. You can share this information with topology-aware applications, such as HDFS, HBase, and Cassandra. These applications use this information to make intelligent data replication decisions for increasing data availability and durability.

If you start or launch an instance in a partition placement group and there is insufficient unique hardware to fulfill the request, the request fails. Amazon EC2 makes more distinct hardware available over time, so you can try your request again later.

### Spread Placement Groups

A spread placement group is a group of instances that are each placed on distinct racks, with each rack having its own network and power source.

The following image shows seven instances in a single Availability Zone that are placed into a spread placement group. The seven instances are placed on seven different racks.

Spread placement groups are recommended for applications that have a small number of critical instances that should be kept separate from each other. Launching instances in a spread placement group reduces the risk of simultaneous failures that might occur when instances share the same racks. Spread placement groups provide access to distinct racks, and are therefore suitable for mixing instance types or launching instances over time.

A spread placement group can span multiple Availability Zones in the same Region. You can have a maximum of seven running instances per Availability Zone per group.

If you start or launch an instance in a spread placement group and there is insufficient unique hardware to fulfill the request, the request fails. Amazon EC2 makes more distinct hardware available over time, so you can try your request again later.
Placement Group Rules and Limitations

General Rules and Limitations

Before you use placement groups, be aware of the following rules:

- The name you specify for a placement group must be unique within your AWS account for the Region.
- You can't merge placement groups.
- An instance can be launched in one placement group at a time; it cannot span multiple placement groups.
- On-Demand Capacity Reservation (p. 318) and zonal Reserved Instances (p. 199) provide a capacity reservation for EC2 instances in a specific Availability Zone. The capacity reservation can be used by instances in a placement group. However, it is not possible to explicitly reserve capacity for a placement group.
- Instances with a tenancy of host cannot be launched in placement groups.

Cluster Placement Group Rules and Limitations

The following rules apply to cluster placement groups:

- The following are the only instance types that you can use when you launch an instance into a cluster placement group:
  - General purpose: M4, M5, M5a, M5ad, M5d
  - Compute optimized: C3, C4, C5, C5d, C5n, and cc2.8xlarge
  - Memory optimized: cr1.8xlarge, R3, R4, R5, R5a, R5ad, R5d, X1, X1e, and z1d
  - Storage optimized: D2, H1, hs1.8xlarge, I2, I3, and I3en
  - Accelerated computing: F1, G2, G3, G4dn, P2, P3, and P3dn
- A cluster placement group can't span multiple Availability Zones.
- The maximum network throughput speed of traffic between two instances in a cluster placement group is limited by the slower of the two instances. For applications with high-throughput requirements, choose an instance type with network connectivity that meets your requirements.
- For instances that are enabled for enhanced networking, the following rules apply:
  - Instances within a cluster placement group can use up to 10 Gbps for single-flow traffic. Instances that are not within a cluster placement group can use up to 5Gbps for single-flow traffic.
  - Traffic to and from Amazon S3 buckets within the same Region over the public IP address space or through a VPC endpoint can use all available instance aggregate bandwidth.
- You can launch multiple instance types into a cluster placement group. However, this reduces the likelihood that the required capacity will be available for your launch to succeed. We recommend using the same instance type for all instances in a cluster placement group.
- Network traffic to the internet and over an AWS Direct Connect connection to on-premises resources is limited to 5 Gbps.

Partition Placement Group Rules and Limitations

The following rules apply to partition placement groups:

- A partition placement group supports a maximum of seven partitions per Availability Zone. The number of instances that you can launch in a partition placement group is limited only by your account limits.
• When instances are launched into a partition placement group, Amazon EC2 tries to evenly distribute
the instances across all partitions. Amazon EC2 doesn’t guarantee an even distribution of instances
across all partitions.
• A partition placement group with Dedicated Instances can have a maximum of two partitions.
• Partition placement groups are not supported for Dedicated Hosts.

Spread Placement Group Rules and Limitations

The following rules apply to spread placement groups:

• A spread placement group supports a maximum of seven running instances per Availability Zone. For
example, in a Region with three Availability Zones, you can run a total of 21 instances in the group
(seven per zone). If you try to start an eighth instance in the same Availability Zone and in the same
spread placement group, the instance will not launch. If you need to have more than seven instances in
an Availability Zone, then the recommendation is to use multiple spread placement groups. This does
not provide guarantees about the spread of instances between groups, but does ensure the spread for
each group to limit impact from certain classes of failures.
• Spread placement groups are not supported for Dedicated Instances or Dedicated Hosts.

Creating a Placement Group

You can create a placement group using the Amazon EC2 console or the command line.

To create a placement group (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Placement Groups, Create Placement Group.
3. Specify a name for the group.
4. Choose the strategy for the group. If you choose Partition, specify the number of partitions within
   the group.
5. Choose Create.

To create a placement group (command line)

• create-placement-group (AWS CLI)
• New-EC2PlacementGroup (AWS Tools for Windows PowerShell)

To create a partition placement group (AWS CLI)

• Use the create-placement-group command and specify the --strategy parameter with the value
  partition and the --partition-count parameter. In this example, the partition placement
group is named HDFS-Group-A and is created with five partitions.

aws ec2 create-placement-group --group-name HDFS-Group-A --strategy partition --
partition-count 5
Launching Instances in a Placement Group

You can create an AMI specifically for the instances to be launched in a placement group. To do this, launch an instance and install the required software and applications on the instance. Then, create an AMI from the instance. For more information, see Creating a Custom Windows AMI (p. 64).

To launch instances into a placement group (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Choose Launch Instance. Complete the wizard as directed, taking care to do the following:

   - On the Choose an Amazon Machine Image (AMI) page, select an AMI. To select an AMI you created, choose My AMIs.
   - On the Choose an Instance Type page, select an instance type that can be launched into a placement group.
   - On the Configure Instance Details page, the following fields are applicable to placement groups:
     - For Number of instances, enter the total number of instances that you need in this placement group, because you might not be able to add instances to the placement group later.
     - For Placement group, select the Add instance to placement group check box. If you do not see Placement group on this page, verify that you have selected an instance type that can be launched into a placement group; otherwise, this option is not available.
     - For Placement group name, you can choose to add the instances to an existing placement group or to a new placement group that you create.
     - For Placement group strategy, choose the appropriate strategy. If you choose partition, for Target partition, choose Auto distribution to have Amazon EC2 do a best effort to distribute the instances evenly across all the partitions in the group; or, specify the partition in which to launch the instances.

To launch instances into a placement group (command line)

1. Create an AMI for your instances using one of the following commands:
   - create-image (AWS CLI)
   - New-EC2Image (AWS Tools for Windows PowerShell)
2. Launch instances into your placement group using one of the following options:
   - --placement with run-instances (AWS CLI)
   - -PlacementGroup with New-EC2Instance (AWS Tools for Windows PowerShell)

To launch instances into a specific partition of a partition placement group (AWS CLI)

- Use the run-instances command and specify the placement group name and partition using the --placement "GroupName = HDFS-Group-A, PartitionNumber = 3" parameter. In this example, the placement group is named HDFS-Group-A and the partition number is 3.

aws ec2 run-instances --placement "GroupName = HDFS-Group-A, PartitionNumber = 3"
Describing Instances in a Placement Group

You can view the placement information of your instances using the Amazon EC2 console or the command line. The placement group is viewable using the console. The partition number for instances in a partition placement group is currently only viewable using the API or AWS CLI.

To view the placement group of an instance (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance and, in the details pane, inspect Placement group. If the instance is not in a placement group, the field is empty. Otherwise, the placement group name is displayed. If the placement group is a partition placement group, inspect Partition number for the partition number for the instance.

To view the partition number for an instance in a partition placement group (AWS CLI)

- Use the describe-instances command and specify the --instance-id parameter.

```sh
aws ec2 describe-instances --instance-id i-0123a456700123456
```

The response contains the placement information, which includes the placement group name and the partition number for the instance.

```
"Placement": {
   "AvailabilityZone": "us-east-1c",
   "GroupName": "HDFS-Group-A",
   "PartitionNumber": 3,
   "Tenancy": "default"
}
```

To filter instances for a specific partition placement group and partition number (AWS CLI)

- Use the describe-instances command and specify the --filters parameter with the placement-group-name and placement-partition-number filters. In this example, the placement group is named HDFS-Group-A and the partition number is 7.

```sh
aws ec2 describe-instances --filters "Name = placement-group-name, Values = HDFS-Group-A" "Name = placement-partition-number, Values = 7"
```

The response lists all the instances that are in the specified partition within the specified placement group. The following is example output showing only the instance ID, instance type, and placement information for the returned instances.

```
"Instances": [ 
   { 
      "InstanceId": "i-0a1bc23d4567e8f90",
      "InstanceType": "r4.large",
   },
   
   "Placement": { 
      "AvailabilityZone": "us-east-1c",
      "GroupName": "HDFS-Group-A",
      "PartitionNumber": 7,
      "Tenancy": "default"
   }
]
Changing the Placement Group for an Instance

You can move an existing instance to a placement group, move an instance from one placement group to another, or remove an instance from a placement group. Before you begin, the instance must be in the stopped state.

You can change the placement group for an instance using the command line or an AWS SDK.

**To move an instance to a placement group (command line)**

1. Stop the instance using one of the following commands:
   - `stop-instances` (AWS CLI)
   - `Stop-EC2Instance` (AWS Tools for Windows PowerShell)
2. Use the `modify-instance-placement` command (AWS CLI) and specify the name of the placement group to which to move the instance.

   ```bash
   aws ec2 modify-instance-placement --instance-id i-0123a456700123456 --group-name MySpreadGroup
   ```

   Alternatively, use the `Edit-EC2InstancePlacement` command (AWS Tools for Windows PowerShell).
3. Restart the instance using one of the following commands:
   - `start-instances` (AWS CLI)
   - `Start-EC2Instance` (AWS Tools for Windows PowerShell)

**To remove an instance from a placement group (command line)**

1. Stop the instance using one of the following commands:
   - `stop-instances` (AWS CLI)
   - `Stop-EC2Instance` (AWS Tools for Windows PowerShell)
2. Use the `modify-instance-placement` command (AWS CLI) and specify an empty string for the group name.

   ```bash
   aws ec2 modify-instance-placement --instance-id i-0123a456700123456 --group-name ""
   ```

   Alternatively, use the `Edit-EC2InstancePlacement` command (AWS Tools for Windows PowerShell).
3. Restart the instance using one of the following commands:
Deleting a Placement Group

If you need to replace a placement group or no longer need one, you can delete it. Before you can delete your placement group, you must terminate all instances that you launched into the placement group, or move them to another placement group.

To terminate or move instances and delete a placement group (console)

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select and terminate all instances in the placement group. You can verify that the instance is in a placement group before you terminate it by checking the value of Placement Group in the details pane.

   Alternatively, follow the steps in Changing the Placement Group for an Instance (p. 717) to move the instances to a different placement group.
4. In the navigation pane, choose Placement Groups.
5. Select the placement group and choose Delete Placement Group.
6. When prompted for confirmation, choose Delete.

To terminate instances and delete a placement group (command line)

You can use one of the following sets of commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- terminate-instances and delete-placement-group (AWS CLI)
- Remove-EC2Instance and Remove-EC2PlacementGroup (AWS Tools for Windows PowerShell)

Network Maximum Transmission Unit (MTU) for Your EC2 Instance

The maximum transmission unit (MTU) of a network connection is the size, in bytes, of the largest permissible packet that can be passed over the connection. The larger the MTU of a connection, the more data that can be passed in a single packet. Ethernet packets consist of the frame, or the actual data you are sending, and the network overhead information that surrounds it.

Ethernet frames can come in different formats, and the most common format is the standard Ethernet v2 frame format. It supports 1500 MTU, which is the largest Ethernet packet size supported over most of the Internet. The maximum supported MTU for an instance depends on its instance type. All Amazon EC2 instance types support 1500 MTU, and many current instance sizes support 9001 MTU, or jumbo frames.

Contents
- Jumbo Frames (9001 MTU) (p. 719)
- Path MTU Discovery (p. 719)
- Check the Path MTU Between Two Hosts (p. 719)
- Check and Set the MTU on Your Windows Instance (p. 720)
Jumbo Frames (9001 MTU)

Jumbo frames allow more than 1500 bytes of data by increasing the payload size per packet, and thus increasing the percentage of the packet that is not packet overhead. Fewer packets are needed to send the same amount of usable data. However, outside of a given AWS region (EC2-Classic), a single VPC, or a VPC peering connection, you will experience a maximum path of 1500 MTU. VPN connections and traffic sent over an Internet gateway are limited to 1500 MTU. If packets are over 1500 bytes, they are fragmented, or they are dropped if the Don’t Fragment flag is set in the IP header.

Jumbo frames should be used with caution for Internet-bound traffic or any traffic that leaves a VPC. Packets are fragmented by intermediate systems, which slows down this traffic. To use jumbo frames inside a VPC and not slow traffic that’s bound for outside the VPC, you can configure the MTU size by route, or use multiple elastic network interfaces with different MTU sizes and different routes.

For instances that are collocated inside a cluster placement group, jumbo frames help to achieve the maximum network throughput possible, and they are recommended in this case. For more information, see Placement Groups (p. 710).

You can use jumbo frames for traffic between your VPCs and your on-premises networks over AWS Direct Connect. For more information, and for how to verify Jumbo Frame capability, see Setting Network MTU in the AWS Direct Connect User Guide.

All current generation instances (p. 123) support jumbo frames. The following previous generation instances support jumbo frames: C3, G2, I2, M3, and R3.

Path MTU Discovery

Path MTU Discovery is used to determine the path MTU between two devices. The path MTU is the maximum packet size that’s supported on the path between the originating host and the receiving host. If a host sends a packet that’s larger than the MTU of the receiving host or that’s larger than the MTU of a device along the path, the receiving host or device returns the following ICMP message: Destination Unreachable: Fragmentation Needed and Don’t Fragment was Set (Type 3, Code 4). This instructs the original host to adjust the MTU until the packet can be transmitted.

By default, security groups do not allow any inbound ICMP traffic. To ensure that your instance can receive this message and the packet does not get dropped, you must add a Custom ICMP Rule with the Destination Unreachable protocol to the inbound security group rules for your instance. For more information, see Rules for Path MTU Discovery (p. 594).

Important

Modifying your instance’s security group to allow path MTU discovery does not guarantee that jumbo frames will not be dropped by some routers. An Internet gateway in your VPC will forward packets up to 1500 bytes only. 1500 MTU packets are recommended for Internet traffic.

Check the Path MTU Between Two Hosts

You can check the path MTU between two hosts using the mturoute.exe command, which you can download and install from http://www.elifulkerson.com/projects/mturoute.php.

To check path MTU using mturoute.exe

2. Open a Command Prompt window and change to the directory where you downloaded mturoute.exe.
3. Use the following command to check the path MTU between your EC2 instance and another host. You can use a DNS name or an IP address as the destination. If the destination is another EC2 instance, verify that the security group allows inbound UDP traffic. This example checks the path MTU between an EC2 instance and www.elifulkerson.com.

```
\mturoute.exe www.elifulkerson.com
* ICMP Fragmentation is not permitted. *
* Speed optimization is enabled. *
* Maximum payload is 10000 bytes. *
+ ICMP payload of 1472 bytes succeeded.
- ICMP payload of 1473 bytes is too big.
Path MTU: 1500 bytes.
```

In this example, the path MTU is 1500.

**Check and Set the MTU on Your Windows Instance**

Some drivers are configured to use jumbo frames, and others are configured to use standard frame sizes. You might want to use jumbo frames for network traffic within your VPC or standard frames for Internet traffic. Whatever your use case, we recommend that you verify that your instances behave as expected.

**ENA Driver**

You can change the MTU setting using Device Manager or the `Set-NetAdapterAdvancedProperty` command.

To get the current MTU setting using the `Get-NetAdapterAdvancedProperty` command, use the following command. Check the entry for the interface name **MTU**. A value of 9001 indicates that Jumbo frames are enabled. Jumbo frames are disabled by default.

```
Get-NetAdapterAdvancedProperty -Name "Ethernet"
```

Enable jumbo frames as follows:

```
Set-NetAdapterAdvancedProperty -Name "Ethernet" -RegistryKeyword "MTU" -RegistryValue 9001
```

Disable jumbo frames as follows:

```
Set-NetAdapterAdvancedProperty -Name "Ethernet" -RegistryKeyword "MTU" -RegistryValue 1500
```

**Intel SRIOV 82599 Driver**

You can change the MTU setting using Device Manager or the `Set-NetAdapterAdvancedProperty` command.

To get the current MTU setting using the `Get-NetAdapterAdvancedProperty` command, use the following command. Check the entry for the interface name **JumboPacket**. A value of 9014 indicates that Jumbo frames are enabled. (Note that the MTU size includes the header and the payload.) Jumbo frames are disabled by default.

```
Get-NetAdapterAdvancedProperty -Name "Ethernet"
```

Enable jumbo frames as follows:
Set-NetAdapterAdvancedProperty -Name "Ethernet" -RegistryKeyword "*JumboPacket" - RegistryValue 9014

Disable jumbo frames as follows:

Set-NetAdapterAdvancedProperty -Name "Ethernet" -RegistryKeyword "*JumboPacket" - RegistryValue 1514

AWS PV Driver

You cannot change the MTU setting using Device Manager, but you can change it using the `netsh` command.

Get the current MTU setting using the following command. The name of the interface can vary. In the output, look for an entry with the name "Ethernet," "Ethernet 2," or "Local Area Connection". You'll need the interface name to enable or disable jumbo frames. A value of 9001 indicates that Jumbo frames are enabled.

```
netsh interface ipv4 show subinterface
```

Enable jumbo frames as follows:

```
netsh interface ipv4 set subinterface "Ethernet" mtu=9001
```

Disable jumbo frames as follows:

```
netsh interface ipv4 set subinterface "Ethernet" mtu=1500
```

Troubleshooting

If you experience connectivity issues between your EC2 instance and an Amazon Redshift cluster when using jumbo frames, see Queries Appear to Hang in the Amazon Redshift Cluster Management Guide.

Virtual Private Clouds

Amazon Virtual Private Cloud (Amazon VPC) enables you to define a virtual network in your own logically isolated area within the AWS cloud, known as a virtual private cloud (VPC). You can launch your Amazon EC2 resources, such as instances, into the subnets of your VPC. Your VPC closely resembles a traditional network that you might operate in your own data center, with the benefits of using scalable infrastructure from AWS. You can configure your VPC; you can select its IP address range, create subnets, and configure route tables, network gateways, and security settings. You can connect instances in your VPC to the internet or to your own data center.

When you create your AWS account, we create a default VPC for you in each region. A default VPC is a VPC that is already configured and ready for you to use. You can launch instances into your default VPC immediately. Alternatively, you can create your own nondefault VPC and configure it as you need.

If you created your AWS account before 2013-12-04, you might have support for the EC2-Classic platform in some regions. If you created your AWS account after 2013-12-04, it does not support EC2-Classic, so you must launch your resources in a VPC. For more information, see EC2-Classic (p. 752).
Amazon VPC Documentation

For more information about Amazon VPC, see the following documentation.

<table>
<thead>
<tr>
<th>Guide</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon VPC User Guide</td>
<td>Describes key concepts and provides instructions for using the features of Amazon VPC.</td>
</tr>
<tr>
<td>Amazon VPC Peering Guide</td>
<td>Describes VPC peering connections and provides instructions for using them.</td>
</tr>
<tr>
<td>Amazon VPC Network Administrator Guide</td>
<td>Helps network administrators configure customer gateways.</td>
</tr>
</tbody>
</table>

Ports and Protocols for Windows Amazon Machine Images (AMIs)

The following tables list the ports, protocols, and directions by workload for Windows Amazon Machine Images.

Contents
- AllJoyn Router (p. 722)
- Cast to Device (p. 723)
- Core Networking (p. 725)
- Delivery Optimization (p. 743)
- Diag Track (p. 744)
- DIAL Protocol Server (p. 744)
- Distributed File System (DFS) Management (p. 744)
- File and Printer Sharing (p. 745)
- File Server Remote Management (p. 747)
- ICMP v4 All (p. 748)
- Multicast (p. 748)
- Remote Desktop (p. 749)
- Windows Device Management (p. 751)
- Windows Firewall Remote Management (p. 751)
- Windows Remote Management (p. 752)

AllJoyn Router

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Description</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2016</td>
<td>AllJoyn Router (TCP-In)</td>
<td>Inbound rule for AllJoyn</td>
<td>Local: 9955</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Description</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Windows Server 2019</td>
<td>AllJoyn Router (TCP-Out)</td>
<td>Outbound rule for AllJoyn Router traffic [TCP]</td>
<td>Local: Any</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>AllJoyn Router (UDP-In)</td>
<td>Inbound rule for AllJoyn Router traffic [UDP]</td>
<td>Local: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>AllJoyn Router (UDP-Out)</td>
<td>Outbound rule for AllJoyn Router traffic [UDP]</td>
<td>Local: Any</td>
<td>UDP</td>
<td>Out</td>
</tr>
</tbody>
</table>

**Cast to Device**

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Description</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2016</td>
<td>Cast to Device functionality</td>
<td>Inbound rule for the Cast to Device functionality to allow use of the Quality Windows Audio Video Experience Service. [TCP 2177]</td>
<td>Local: 2177</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2019</td>
<td>Cast to Device functionality</td>
<td>Outbound rule for the Cast to Device functionality to allow use of the Quality Windows Audio Video Experience Service. [TCP 2177]</td>
<td>Local: Any</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>Cast to Device functionality</td>
<td>Inbound rule for the Cast to Device functionality to allow use of the Quality Windows Audio Video</td>
<td>Local: 2177</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Description</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>-------------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Cast to Device functionality (qWave-UDP-Out)</td>
<td>Outbound rule for the Cast to Device functionality to allow use of the Quality Windows Audio Video Experience Service. [UDP 2177]</td>
<td>Local: Any Remote: 2177</td>
<td>UDP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>Cast to Device SSDP Discovery (UDP-In)</td>
<td>Inbound rule to allow discovery of Cast to Device targets using SSDP</td>
<td>Local: Ply2Disc Remote: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Cast to Device Streaming Server (HTTP-Streaming-In)</td>
<td>Inbound rule for the Cast to Device server to allow streaming using HTTP. [TCP 10246]</td>
<td>Local: 10246 Remote: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Cast to Device Streaming Server (RTCP-Streaming-In)</td>
<td>Inbound rule for the Cast to Device server to allow streaming using RTSP and RTP. [UDP]</td>
<td>Local: Any Remote: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Cast to Device Streaming Server (RTP-Streaming-Out)</td>
<td>Outbound rule for the Cast to Device server to allow streaming using RTSP and RTP. [UDP]</td>
<td>Local: Any Remote: Any</td>
<td>UDP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>Cast to Device Streaming Server (RTSP-Streaming-In)</td>
<td>Inbound rule for the Cast to Device server to allow streaming using RTSP and RTP. [TCP 23554, 23555, 23556]</td>
<td>Local: 235, 542, 355, 523, 556 Remote: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
</tbody>
</table>
### Core Networking


<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2012</td>
<td>Destination Unreachable (ICMPv6-In)</td>
<td>Destination Unreachable error messages are sent from any node that a packet traverses which is unable to forward the packet for any reason except congestion.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2012 R2</td>
<td>Destination Unreachable Fragmentation Needed (ICMPv4-In)</td>
<td>Destination Unreachable Fragmentation Needed error messages are sent from any node that a packet traverses which is unable to forward the packet because fragmentation was needed and the don't fragment bit was set.</td>
<td>Local: 68</td>
<td>ICMPv4</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2016</td>
<td>Core Networking - DNS (UDP-Out)</td>
<td>Outbound rule to allow DNS requests. DNS responses based on</td>
<td>Local: Any</td>
<td>UDP</td>
<td>Out</td>
</tr>
<tr>
<td>Windows Server 2019</td>
<td>Cast to Device UPnP Events (TCP-In)</td>
<td>Inbound rule to allow receiving UPnP Events from Cast to Device targets</td>
<td>Local: 2869</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>requests that match this rule are permitted regardless of source address. This behavior is classified as loose source mapping.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Host Configuration</td>
<td>Allows DHCP (Dynamic Host</td>
<td>Local: 68 Remote: 67</td>
<td>UDP</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>Protocol (DHCP-In)</td>
<td>Configuration Protocol</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>messages for stateful auto-configuration.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allows DHCP (Dynamic Host</td>
<td>Local: 68 Remote: 67</td>
<td>UDP</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>Dynamic Host Configuration</td>
<td>Configuration Protocol (DHCP-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol (DHCP-Out)</td>
<td>Out)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allows DHCPV6 (Dynamic Host</td>
<td>Local: 546 Remote: 547</td>
<td>UDP</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>Dynamic Host Configuration</td>
<td>Configuration Protocol for</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol for IPv6 (DHCPV6-</td>
<td>IPv6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In)</td>
<td>Allows DHCPV6 (Dynamic Host</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Configuration Protocol for IPv6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic Host Configuration</td>
<td>Protocol for IPv6 (DHCPV6-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol for IPv6 (DHCPV6-</td>
<td>Out)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core Networking -</td>
<td>Outbound rule to allow</td>
<td></td>
<td>TCP</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>Group Policy (LSASS-Out)</td>
<td>remote LSASS traffic for</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group Policy updates.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Core Networking - Group Policy (TCP-Out)</td>
<td>Outbound rule to allow remote RPC traffic for Group Policy updates.</td>
<td>Local: Any Remote: Any</td>
<td>TCP</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>Internet Group Management Protocol (IGMP-In)</td>
<td>IGMP messages are sent and received by nodes to create, join, and depart multicast groups.</td>
<td>Local: 68 Remote: 67</td>
<td>2</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>Core Networking - Internet Group Management Protocol (IGMP-Out)</td>
<td>IGMP messages are sent and received by nodes to create, join, and depart multicast groups.</td>
<td>Local: 68 Remote: 67</td>
<td>2</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>Core Networking - IPHTTPS (TCP-In)</td>
<td>Inbound TCP rule to allow IPHTTPS tunneling technology to provide connectivity across HTTP proxies and firewalls.</td>
<td>Local: IPHTTPS Remote: Any</td>
<td>TCP</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>Core Networking - IPHTTPS (TCP-Out)</td>
<td>Outbound TCP rule to allow IPHTTPS tunneling technology to provide connectivity across HTTP proxies and firewalls.</td>
<td>Local: Any Remote: IPHTTPS</td>
<td>TCP</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>IPv6 (IPv6-In)</td>
<td>Inbound rule required to permit IPv6 traffic for ISATAP (Intra-Site Automatic Tunnel Addressing Protocol) and 6to4 tunneling services.</td>
<td>Local: Any Remote: 445</td>
<td>IPv6</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>IPv6 (IPv6-Out)</td>
<td>Outbound rule required to permit IPv6 traffic for ISATAP (Intra-Site Automatic Tunnel Addressing Protocol) and 6to4 tunneling services.</td>
<td>Local: Any Remote: 445</td>
<td>IPv6</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>Multicast Listener Done (ICMPv6-In)</td>
<td>Multicast Listener Done messages inform local routers that there are no longer any members remaining for a specific multicast address on the subnet.</td>
<td>Local: 68 Remote: 67</td>
<td>ICMPv6</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>Multicast Listener Done (ICMPv6-Out)</td>
<td>Multicast Listener Done messages inform local routers that there are no longer any members remaining for a specific multicast address on the subnet.</td>
<td>Local: 68 Remote: 67</td>
<td>ICMPv6</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Multicast Listener Query (ICMPv6-In)</td>
<td>An IPv6 multicast-capable router uses the Multicast Listener Query message to query a link for multicast group membership.</td>
<td>Local: 68  Remote: 67</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Multicast Listener Query (ICMPv6-Out)</td>
<td>An IPv6 multicast-capable router uses the Multicast Listener Query message to query a link for multicast group membership.</td>
<td>Local: 68  Remote: 67</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>Multicast Listener Report (ICMPv6-In)</td>
<td>The Multicast Listener Report message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address or in response to a Multicast Listener Query.</td>
<td>Local: 68  Remote: 67</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Multicast Listener Report (ICMPv6-Out)</td>
<td>The Multicast Listener Report message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address or in response to a Multicast Listener Query.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multicast Listener Report v2 (ICMPv6-In)</td>
<td>Multicast Listener Report v2 message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address or in response to a Multicast Listener Query.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

730
<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multicast Listener Report v2 (ICMPv6-Out)</td>
<td>Multicast Listener Report v2 message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address or in response to a Multicast Listener Query.</td>
<td>Local: 68 Remote: 67</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Advertisement (ICMPv6-In)</td>
<td>Neighbor Discovery Advertisement messages are sent by nodes to notify other nodes of link-layer address changes or in response to a Neighbor Discovery Solicitation request.</td>
<td>Local: 68 Remote: 67</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Advertisement (ICMPv6-Out)</td>
<td>Neighbor Discovery Advertisement messages are sent by nodes to notify other nodes of link-layer address changes or in response to a Neighbor Discovery Solicitation request.</td>
<td>Local: 68 Remote: 67</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Solicitation (ICMPv6-In)</td>
<td>Neighbor Discovery Solicitations are sent by nodes to discover the link-layer address of another on-link IPv6 node.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Solicitation (ICMPv6-Out)</td>
<td>Neighbor Discovery Solicitations are sent by nodes to discover the link-layer address of another on-link IPv6 node.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Packet Too Big (ICMPv6-In)</td>
<td>Packet Too Big error messages are sent from any node that a packet traverses which is unable to forward the packet because the packet is too large for the next link.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Packet Too Big (ICMPv6-Out)</td>
<td>Packet Too Big error messages are sent from any node that a packet traverses which is unable to forward the packet because the packet is too large for the next link.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Parameter Problem</td>
<td>Parameter Problem (ICMPv6-In)</td>
<td>Parameter Problem error messages are sent by nodes when packets are incorrectly generated.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td>Problem (ICMPv6-Out)</td>
<td>Parameter Problem (ICMPv6-Out)</td>
<td>Parameter Problem error messages are sent by nodes when packets are incorrectly generated.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td>Router Advertisement (ICMPv6-In)</td>
<td>Router Advertisement (ICMPv6-In)</td>
<td>Router Advertisement messages are sent by routers to other nodes for stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td>Router Advertisement (ICMPv6-Out)</td>
<td>Router Advertisement (ICMPv6-Out)</td>
<td>Router Advertisement messages are sent by routers to other nodes for stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td>Router Solicitation (ICMPv6-In)</td>
<td>Router Solicitation (ICMPv6-In)</td>
<td>Router Solicitation messages are sent by nodes seeking routers to provide stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td>Router Solicitation (ICMPv6-Out)</td>
<td>Router Solicitation (ICMPv6-Out)</td>
<td>Router Solicitation messages are sent by nodes seeking routers to provide stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Core Networking - Teredo (UDP-In)</td>
<td>Inbound UDP rule to allow Teredo edge traversal. This technology provides address assignment and automatic tunneling for unicast IPv6 traffic when an IPv6/IPv4 host is located behind an IPv4 network address translator.</td>
<td>Local: Teredo Remote: Any</td>
<td>UDP</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>Core Networking - Teredo (UDP-Out)</td>
<td>Outbound UDP rule to allow Teredo edge traversal. This technology provides address assignment and automatic tunneling for unicast IPv6 traffic when an IPv6/IPv4 host is located behind an IPv4 network address translator.</td>
<td>Local: Any Remote: Any</td>
<td>UDP</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>Time Exceeded (ICMPv6-In)</td>
<td>Time Exceeded error messages are generated from any node that a packet traverses if the Hop Limit value is decremented to zero at any point on the path.</td>
<td>Local: 68 Remote: 67</td>
<td>ICMPv6</td>
<td>In</td>
<td></td>
</tr>
</tbody>
</table>
Time Exceeded (ICMPv6-Out)

Time Exceeded error messages are generated from any node that a packet traverses if the Hop Limit value is decremented to zero at any point on the path.

*OS* | **Rule** | **Definition** | **Port** | **Protocol** | **Direction**
--- | --- | --- | --- | --- | ---
 | Time Exceeded (ICMPv6-Out) | Time Exceeded error messages are generated from any node that a packet traverses if the Hop Limit value is decremented to zero at any point on the path. | Local: 68 Remote: 67 | ICMPv6 | Out

Windows Server 2008 R2 and SP2

Destination Unreachable (ICMPv6-In)

Destination Unreachable error messages are sent from any node that a packet traverses which is unable to forward the packet for any reason except congestion.

*OS* | **Rule** | **Definition** | **Port** | **Protocol** | **Direction**
--- | --- | --- | --- | --- | ---
 | Windows Server 2008 R2 | Destination Unreachable (ICMPv6-In) | Local: 68 Remote: 67 | ICMPv6 | Out
 | Windows Server 2008 SP2 | | | | |

Destination Unreachable Fragmentation Needed (ICMPv4-In)

Destination Unreachable Fragmentation Needed error messages are sent from any node that a packet traverses which is unable to forward the packet because fragmentation was needed and the don't fragment bit was set.

*OS* | **Rule** | **Definition** | **Port** | **Protocol** | **Direction**
--- | --- | --- | --- | --- | ---
 | Destination Unreachable Fragmentation Needed (ICMPv4-In) | Destination Unreachable Fragmentation Needed error messages are sent from any node that a packet traverses which is unable to forward the packet because fragmentation was needed and the don't fragment bit was set. | Local: 68 Remote: 67 | ICMPv4 | In
<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dynamic Host Configuration Protocol (DHCP-In)</td>
<td>Allows DHCP (Dynamic Host Configuration Protocol) messages for stateful auto-configuration.</td>
<td>Local: 68 Remote: 67</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Dynamic Host Configuration Protocol for IPv6 (DHCPV6-In)</td>
<td>Allows DHCPV6 (Dynamic Host Configuration Protocol for IPv6) messages for stateful and stateless configuration.</td>
<td>Local: 546 Remote: 547</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Internet Group Management Protocol (IGMP-In)</td>
<td>IGMP messages are sent and received by nodes to create, join, and depart multicast groups.</td>
<td>Local: 68 Remote: 67</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
</tbody>
</table>
| IPv6 (IPv6-In) | Inbound rule required to permit IPv6 traffic for ISATAP (Intra-Site Automatic Tunnel Addressing Protocol) and 6to4 tunneling services. | Local: Any  
Remote: 445 | 41         | In       |
| IPv6 (IPv6-Out) | Outbound rule required to permit IPv6 traffic for ISATAP (Intra-Site Automatic Tunnel Addressing Protocol) and 6to4 tunneling services. | Local: Any  
Remote: 445 | 41         | Out      |
| Multicast Listener Done (ICMPv6-In) | Multicast Listener Done messages inform local routers that there are no longer any members remaining for a specific multicast address on the subnet. | Local: 68  
Remote: 67 | ICMPv6     | In       |
| Multicast Listener Done (ICMPv6-Out) | Multicast Listener Done messages inform local routers that there are no longer any members remaining for a specific multicast address on the subnet. | Local: 68  
Remote: 67 | ICMPv6     | Out      |
<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
</table>
|    | Multicast Listener Query (ICMPv6-In) | An IPv6 multicast-capable router uses the Multicast Listener Query message to query a link for multicast group membership. | Local: 68  
Remote: 67 | ICMPv6 | In |
|    | Multicast Listener Query (ICMPv6-Out) | An IPv6 multicast-capable router uses the Multicast Listener Query message to query a link for multicast group membership. | Local: 68  
Remote: 67 | ICMPv6 | Out |
|    | Multicast Listener Report (ICMPv6-In) | The Multicast Listener Report message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address, or in response to a Multicast Listener Query. | Local: 68  
Remote: 67 | ICMPv6 | In |
<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multicast Listener Report (ICMPv6-Out)</td>
<td>The Multicast Listener Report message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address, or in response to a Multicast Listener Query.</td>
<td>Local: 68, Remote: 67</td>
<td>ICMPv6</td>
<td>Out</td>
<td></td>
</tr>
<tr>
<td>Multicast Listener Report v2 (ICMPv6-In)</td>
<td>Multicast Listener Report v2 message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address, or in response to a Multicast Listener Query.</td>
<td>Local: 68, Remote: 67</td>
<td>ICMPv6</td>
<td>In</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Multicast Listener Report v2 (ICMPv6-Out)</td>
<td>Multicast Listener Report v2 message is used by a listening node to either immediately report its interest in receiving multicast traffic at a specific multicast address, or in response to a Multicast Listener Query.</td>
<td>Local: 68&lt;br&gt;Remote: 67</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Advertisement (ICMPv6-In)</td>
<td>Neighbor Discovery Advertisement messages are sent by nodes to notify other nodes of link-layer address changes or in response to a Neighbor Discovery Solicitation request.</td>
<td>Local: 68&lt;br&gt;Remote: 67</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Advertisement (ICMPv6-Out)</td>
<td>Neighbor Discovery Advertisement messages are sent by nodes to notify other nodes of link-layer address changes or in response to a Neighbor Discovery Solicitation request.</td>
<td>Local: 68&lt;br&gt;Remote: 67</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Solicitation (ICMPv6-In)</td>
<td>Neighbor Discovery Solicitations are sent by nodes to discover the link-layer address of another on-link IPv6 node.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local: 68  Remote: 67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighbor Discovery Solicitation (ICMPv6-Out)</td>
<td>Neighbor Discovery Solicitations are sent by nodes to discover the link-layer address of another on-link IPv6 node.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local: 68  Remote: 67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Packet Too Big (ICMPv6-In)</td>
<td>Packet Too Big error messages are sent from any node that a packet traverses which is unable to forward the packet because the packet is too large for the next link.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local: 68  Remote: 67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Packet Too Big (ICMPv6-Out)</td>
<td>Packet Too Big error messages are sent from any node that a packet traverses which is unable to forward the packet because the packet is too large for the next link.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local: 68  Remote: 67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Parameter Problem</td>
<td>Parameter Problem (ICMPv6-In)</td>
<td>Parameter Problem error messages are sent by nodes when packets are incorrectly generated.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameter Problem</td>
<td>Parameter Problem (ICMPv6-Out)</td>
<td>Parameter Problem error messages are sent by nodes when packets are incorrectly generated.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router Advertisement</td>
<td>Router Advertisement (ICMPv6-In)</td>
<td>Router Advertisement messages are sent by routers to other nodes for stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router Advertisement</td>
<td>Router Advertisement (ICMPv6-Out)</td>
<td>Router Advertisement messages are sent by routers to other nodes for stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router Solicitation</td>
<td>Router Solicitation (ICMPv6-In)</td>
<td>Router Solicitation messages are sent by nodes seeking routers to provide stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Router Solicitation</td>
<td>Router Solicitation (ICMPv6-Out)</td>
<td>Router Solicitation messages are sent by nodes seeking routers to provide stateless auto-configuration.</td>
<td>Local: 68</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Delivery Optimization

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DeliveryOptimization-</td>
<td>rule to allow Delivery Optimization to connect to remote endpoints.</td>
<td>Local: 7680</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>TCP-In</td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DeliveryOptimization-</td>
<td>rule to allow Delivery Optimization to connect to remote endpoints.</td>
<td>Local: 7680</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>UDP-In</td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Time Exceeded (ICMPv6-In)**
  - Time Exceeded error messages are generated from any node that a packet traverses if the Hop Limit value is decremented to zero at any point on the path.
  - Local: 68
  - Remote: 67
  - Protocol: ICMPv6
  - Direction: In

- **Time Exceeded (ICMPv6-Out)**
  - Time Exceeded error messages are generated from any node that a packet traverses if the Hop Limit value is decremented to zero at any point on the path.
  - Local: 68
  - Remote: 67
  - Protocol: ICMPv6
  - Direction: Out
# Diag Track

## Windows Server 2019

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 443</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Windows Server 2016

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2016</td>
<td>Connected User Experiences and Telemetry</td>
<td>Unified Telemetry Client Outbound Traffic</td>
<td>Local: Any</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# DIAL Protocol Server

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2016</td>
<td>DIAL protocol server (HTTP-In)</td>
<td>Inbound rule for DIAL protocol server to allow</td>
<td>Local: 10247</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td>remote control of Apps using HTTP.</td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Distributed File System (DFS) Management

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2008 R2</td>
<td>DFS Management (SMB-In)</td>
<td>Inbound rule to allow SMB traffic to manage the</td>
<td>Local: 445</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td>File Services role.</td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFS Management (WMI-In)</td>
<td>DFS Management (WMI-In)</td>
<td>Inbound rule to allow WMI traffic to manage the</td>
<td>Local: RPC</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
File and Printer Sharing

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2008 R2</td>
<td>File and Printer Sharing (Echo Request - ICMPv4-In)</td>
<td>Echo Request messages are sent as ping requests to other nodes.</td>
<td>Local: 5355</td>
<td>ICMPv4</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2008 SP2</td>
<td>File and Printer Sharing (Echo Request - ICMPv4-Out)</td>
<td>Echo Request messages are sent as ping requests to other nodes.</td>
<td>Local: 5355</td>
<td>ICMPv4</td>
<td>Out</td>
</tr>
<tr>
<td>Windows Server 2012</td>
<td>File and Printer Sharing (Echo Request - ICMPv6-In)</td>
<td>Echo Request messages are sent as ping requests to other nodes.</td>
<td>Local: 5355</td>
<td>ICMPv6</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2012 R2</td>
<td>File and Printer Sharing (Echo Request - ICMPv6-Out)</td>
<td>Echo Request messages are sent as ping requests to other nodes.</td>
<td>Local: 5355</td>
<td>ICMPv6</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (LLMNR-UDP-In)</td>
<td>Inbound rule for File and Printer Sharing to allow Link Local Multicast Name Resolution.</td>
<td>Local: 5355</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (LLMNR-UDP-Out)</td>
<td>Outbound rule for File and Printer Sharing to allow Link Local Multicast Name Resolution.</td>
<td>Local: Any Remote: 5355</td>
<td>UDP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (NB-Datagram-In)</td>
<td>Inbound rule for File and Printer Sharing to allow NetBIOS Datagram transmission and reception.</td>
<td>Local: 138 Remote: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (NB-Datagram-Out)</td>
<td>Outbound rule for File and Printer Sharing to allow NetBIOS Datagram transmission and reception.</td>
<td>Local: Any Remote: 138</td>
<td>UDP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (NB-Name-In)</td>
<td>Inbound rule for File and Printer Sharing to allow NetBIOS Name Resolution.</td>
<td>Local: 137 Remote: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (NB-Name-Out)</td>
<td>Outbound rule for File and Printer Sharing to allow NetBIOS Name Resolution.</td>
<td>Local: Any Remote: 137</td>
<td>UDP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (NB-Session-In)</td>
<td>Inbound rule for File and Printer Sharing to allow NetBIOS Session Service connections.</td>
<td>Local: 139 Remote: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (NB-Session-Out)</td>
<td>Outbound rule for File and Printer Sharing to allow NetBIOS Session Service connections.</td>
<td>Local: Any Remote: 139</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td>OS</td>
<td>Rule</td>
<td>Definition</td>
<td>Port</td>
<td>Protocol</td>
<td>Direction</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (SMB-In)</td>
<td>Inbound rule for File and Printer Sharing to allow Server Message Block transmission and reception via Named Pipes.</td>
<td>Local: 445</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (SMB-Out)</td>
<td>Outbound rule for File and Printer Sharing to allow Server Message Block transmission and reception via Named Pipes.</td>
<td>Local: Any</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: 445</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (Spooler Service - RPC)</td>
<td>Inbound rule for File and Printer Sharing to allow the Print Spooler Service to communicate via TCP/RPC.</td>
<td>Local: RPC</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>File and Printer Sharing (Spooler Service - RPC-EPMap)</td>
<td>Inbound rule for the RPCSS service to allow RPC/TCP traffic for the Spooler Service.</td>
<td>Local: RPC-EPMap</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### File Server Remote Management

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2008 SP2</td>
<td>File Server Remote Management (DCOM-In)</td>
<td>Inbound rule to allow DCOM traffic to manage the File Services role.</td>
<td>Local: 135</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2012</td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2012 R2</td>
<td>File Server Remote Management (SMB-In)</td>
<td>Inbound rule to allow SMB traffic to manage the File Services role.</td>
<td>Local: 445</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### ICMP v4 All

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2012</td>
<td>All ICMP v4</td>
<td>Inbound rule to allow WMI traffic to manage the File Services role.</td>
<td>Local: 139, Remote: Any</td>
<td>ICMPv4</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2012 R2</td>
<td>All ICMP v4</td>
<td>Inbound rule to allow WMI traffic to manage the File Services role.</td>
<td>Local: 139, Remote: Any</td>
<td>ICMPv4</td>
<td>In</td>
</tr>
</tbody>
</table>

### Multicast

**Windows Server 2019**

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2019</td>
<td>mDNS (UDP-In)</td>
<td>Inbound rule for mDNS traffic.</td>
<td>Local: 5353, Remote: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>mDNS (UDP-Out)</td>
<td>Outbound rule for mDNS traffic.</td>
<td>Local: Any, Remote: 5353</td>
<td>UDP</td>
<td>Out</td>
</tr>
</tbody>
</table>

**Windows Server 2016**

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2016</td>
<td>mDNS (UDP-In)</td>
<td>Inbound rule for mDNS traffic.</td>
<td>Local: mDNS, Remote: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>mDNS (UDP-Out)</td>
<td>Outbound rule for mDNS traffic.</td>
<td>Local: 5353, Remote: Any</td>
<td>UDP</td>
<td>Out</td>
</tr>
</tbody>
</table>
## Remote Desktop

Windows Server 2012 R2, 2016, and 2019

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2012 R2</td>
<td>Remote Desktop - Shadow (TCP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow shadowing of an existing Remote Desktop session.</td>
<td>Local: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2016</td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remote Desktop - User Mode (TCP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow RDP traffic.</td>
<td>Local: 3389</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remote Desktop - User Mode (UDP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow RDP traffic.</td>
<td>Local: 3389</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Windows Server 2012

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2012</td>
<td>Remote Desktop - User Mode (TCP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow RDP traffic.</td>
<td>Local: 3389</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remote Desktop - User Mode (UDP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow RDP traffic.</td>
<td>Local: 3389</td>
<td>UDP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Windows Server 2008 SP2

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2008 SP2</td>
<td>Remote Desktop - Shadow (TCP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow shadowing of an existing Remote Desktop session.</td>
<td>Local: Any, Remote: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remote Desktop - User Mode (TCP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow RDP traffic.</td>
<td>Local: 3389, Remote: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Remote Desktop - User Mode (UDP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow RDP traffic.</td>
<td>Local: 3389, Remote: Any</td>
<td>UDP</td>
<td>In</td>
</tr>
</tbody>
</table>

Windows Server 2008 R2

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2008 R2</td>
<td>RemoteFX (TCP-In)</td>
<td>Inbound rule for the Remote Desktop service to allow RDP traffic.</td>
<td>Local: 3389, Remote: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td>TCP-In</td>
<td></td>
<td></td>
<td>Local: 3389, Remote: Any</td>
<td>TCP</td>
<td>In</td>
</tr>
</tbody>
</table>
## Windows Device Management

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Windows Device Management Enrollment Service (TCP out)</td>
<td>Allow outbound TCP traffic from Windows Device Management Enrollment Service.</td>
<td>Local: Any</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Windows Device Management Sync Client (TCP out)</td>
<td>Allow outbound TCP traffic from Windows Device Management Sync Client.</td>
<td>Local: Any</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Windows Enrollment WinRT (TCP Out)</td>
<td>Allow outbound TCP traffic from Windows Enrollment WinRT.</td>
<td>Local: Any</td>
<td>TCP</td>
<td>Out</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Windows Firewall Remote Management

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2008 SP2</td>
<td>Windows Firewall Remote Management (RPC)</td>
<td>Inbound rule for the Windows Firewall to be remotely managed via RPC/TCP.</td>
<td>Local: RPC</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2012 R2</td>
<td>Windows Firewall Remote Management (RPC-EPMAP)</td>
<td>Inbound rule for the RPCSS service to allow RPC/TCP traffic for RPC-EPMap</td>
<td>Local: RPC-EPMap</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Windows Remote Management

<table>
<thead>
<tr>
<th>OS</th>
<th>Rule</th>
<th>Definition</th>
<th>Port</th>
<th>Protocol</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Server 2008 R2</td>
<td>Windows Remote Management (HTTP-In)</td>
<td>Inbound rule for Windows Remote Management via WS-Management.</td>
<td>Local: 5985</td>
<td>TCP</td>
<td>In</td>
</tr>
<tr>
<td>Windows Server 2008 SP2</td>
<td></td>
<td></td>
<td>Remote: Any</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2012 R2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows Server 2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For more information about Amazon EC2 security groups, see Amazon EC2 Security Groups for Windows Instances.

EC2-Classic

With EC2-Classic, your instances run in a single, flat network that you share with other customers. With Amazon VPC, your instances run in a virtual private cloud (VPC) that’s logically isolated to your AWS account.

The EC2-Classic platform was introduced in the original release of Amazon EC2. If you created your AWS account after 2013-12-04, it does not support EC2-Classic, so you must launch your Amazon EC2 instances in a VPC.

If your account does not support EC2-Classic, we create a default VPC for you. By default, when you launch an instance, we launch it into your default VPC. Alternatively, you can create a nondefault VPC and specify it when you launch an instance.

Detecting Supported Platforms

The Amazon EC2 console indicates which platforms you can launch instances into for the selected region, and whether you have a default VPC in that region.

Verify that the region you’ll use is selected in the navigation bar. On the Amazon EC2 console dashboard, look for Supported Platforms under Account Attributes.
Accounts that Support EC2-Classic

The dashboard displays the following under Account Attributes to indicate that the account supports both the EC2-Classic platform and VPCs in this region, but the region does not have a default VPC.

```
Account Attributes

Supported Platforms
EC2
VPC
```

The output of the describe-account-attributes command includes both the EC2 and VPC values for the supported-platforms attribute.

```
aws ec2 describe-account-attributes --attribute-names supported-platforms
{
  "AccountAttributes": [
    {
      "AttributeName": "supported-platforms",
      "AttributeValues": [
        {
          "AttributeValue": "EC2"
        },
        {
          "AttributeValue": "VPC"
        }
      ]
    }
  ]
}
```

Accounts that Require a VPC

The dashboard displays the following under Account Attributes to indicate that the account requires a VPC to launch instances in this region, does not support the EC2-Classic platform in this region, and the region has a default VPC with the identifier vpc-1a2b3c4d.

```
Account Attributes

Supported Platforms
VPC
Default VPC
vpc-1a2b3c4d
```

The output of the describe-account-attributes command includes only the VPC value for the supported-platforms attribute.

```
aws ec2 describe-account-attributes --attribute-names supported-platforms
{
  "AccountAttributes": [
    {
      "AttributeName": "supported-platforms",
      "AttributeValues": [
        {
          "AttributeValue": "VPC"
        }
      ]
    }
  ]
}
```
Instance Types Available in EC2-Classic

Most of the newer instance types require a VPC. The following are the only instance types supported in EC2-Classic:

- General purpose: M1, M3, and T1
- Compute optimized: C1, C3, and CC2
- Memory optimized: CR1, M2, and R3
- Storage optimized: D2, HS1, and I2
- Accelerated computing: G2

If your account supports EC2-Classic but you have not created a nondefault VPC, you can do one of the following to launch instances that require a VPC:

- Create a nondefault VPC and launch your VPC-only instance into it by specifying a subnet ID or a network interface ID in the request. Note that you must create a nondefault VPC if you do not have a default VPC and you are using the AWS CLI, Amazon EC2 API, or AWS SDK to launch a VPC-only instance. For more information, see Create a Virtual Private Cloud (VPC) (p. 16).
- Launch your VPC-only instance using the Amazon EC2 console. The Amazon EC2 console creates a nondefault VPC in your account and launches the instance into the subnet in the first Availability Zone. The console creates the VPC with the following attributes:
  - One subnet in each Availability Zone, with the public IPv4 addressing attribute set to true so that instances receive a public IPv4 address. For more information, see IP Addressing in Your VPC in the Amazon VPC User Guide.
  - An Internet gateway, and a main route table that routes traffic in the VPC to the Internet gateway. This enables the instances you launch in the VPC to communicate over the Internet. For more information, see Internet Gateways in the Amazon VPC User Guide.
  - A default security group for the VPC and a default network ACL that is associated with each subnet. For more information, see Security in Your VPC in the Amazon VPC User Guide.

If you have other resources in EC2-Classic, you can take steps to migrate them to a VPC. For more information, see Migrating from a Windows Instance in EC2-Classic to a Windows Instance in a VPC (p. 772).

Differences Between Instances in EC2-Classic and a VPC

The following table summarizes the differences between instances launched in EC2-Classic, instances launched in a default VPC, and instances launched in a nondefault VPC.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>EC2-Classic</th>
<th>Default VPC</th>
<th>Nondefault VPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public IPv4 address (from Amazon's</td>
<td>Your instance receives a public IPv4</td>
<td>Your instance launched in a default subnet receives a public IPv4 address</td>
<td>Your instance doesn't receive a public IPv4 address by default, unless you</td>
</tr>
<tr>
<td>public IP address pool)</td>
<td>address from the EC2-Classic public IPv4 address pool.</td>
<td>by default, unless you specify otherwise during launch, or you modify the subnet's public IPv4 address attribute.</td>
<td>specify otherwise during launch, or you modify the subnet's public IPv4 address attribute.</td>
</tr>
<tr>
<td>Characteristic</td>
<td>EC2-Classic</td>
<td>Default VPC</td>
<td>Nondefault VPC</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Private IPv4 address</td>
<td>Your instance receives a private IPv4 address from the EC2-Classic range each time it's started.</td>
<td>Your instance receives a static private IPv4 address from the address range of your default VPC.</td>
<td>Your instance receives a static private IPv4 address from the address range of your VPC.</td>
</tr>
<tr>
<td>Multiple private IPv4 addresses</td>
<td>We select a single private IP address for your instance; multiple IP addresses are not supported.</td>
<td>You can assign multiple private IPv4 addresses to your instance.</td>
<td>You can assign multiple private IPv4 addresses to your instance.</td>
</tr>
<tr>
<td>Elastic IP address (IPv4)</td>
<td>An Elastic IP is disassociated from your instance when you stop it.</td>
<td>An Elastic IP remains associated with your instance when you stop it.</td>
<td>An Elastic IP remains associated with your instance when you stop it.</td>
</tr>
<tr>
<td>Associating an Elastic IP address</td>
<td>You associate an Elastic IP address with an instance.</td>
<td>An Elastic IP address is a property of a network interface. You associate an Elastic IP address with an instance by updating the network interface attached to the instance.</td>
<td>An Elastic IP address is a property of a network interface. You associate an Elastic IP address with an instance by updating the network interface attached to the instance.</td>
</tr>
<tr>
<td>Reassociating an Elastic IP address</td>
<td>If the Elastic IP address is already associated with another instance, the address is automatically associated with the new instance.</td>
<td>If the Elastic IP address is already associated with another instance, the address is automatically associated with the new instance.</td>
<td>If the Elastic IP address is already associated with another instance, it succeeds only if you allowed reassociation.</td>
</tr>
<tr>
<td>Tagging Elastic IP addresses</td>
<td>You cannot apply tags to an Elastic IP address.</td>
<td>You can apply tags to an Elastic IP address.</td>
<td>You can apply tags to an Elastic IP address.</td>
</tr>
<tr>
<td>DNS hostnames</td>
<td>DNS hostnames are enabled by default.</td>
<td>DNS hostnames are enabled by default.</td>
<td>DNS hostnames are disabled by default.</td>
</tr>
<tr>
<td>Security group</td>
<td>A security group can reference security groups that belong to other AWS accounts.</td>
<td>A security group can reference security groups for your VPC, or for a peer VPC in a VPC peering connection.</td>
<td>A security group can reference security groups for your VPC only.</td>
</tr>
</tbody>
</table>
### Differences Between Instances in EC2-Classic and a VPC

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>EC2-Classic</th>
<th>Default VPC</th>
<th>Nondefault VPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security group association</td>
<td>You can assign an unlimited number of security groups to an instance when you launch it. You can't change the security groups of your running instance. You can either modify the rules of the assigned security groups, or replace the instance with a new one (create an AMI from the instance, launch a new instance from this AMI with the security groups that you need, disassociate any Elastic IP address from the original instance and associate it with the new instance, and then terminate the original instance).</td>
<td>You can assign up to 5 security groups to an instance. You can assign security groups to your instance when you launch it and while it's running.</td>
<td>You can assign up to 5 security groups to an instance. You can assign security groups to your instance when you launch it and while it's running.</td>
</tr>
<tr>
<td>Security group rules</td>
<td>You can add rules for inbound traffic only.</td>
<td>You can add rules for inbound and outbound traffic.</td>
<td>You can add rules for inbound and outbound traffic.</td>
</tr>
<tr>
<td>Tenancy</td>
<td>Your instance runs on shared hardware.</td>
<td>You can run your instance on shared hardware or single-tenant hardware.</td>
<td>You can run your instance on shared hardware or single-tenant hardware.</td>
</tr>
<tr>
<td>Accessing the Internet</td>
<td>Your instance can access the Internet. Your instance automatically receives a public IP address, and can access the Internet directly through the AWS network edge.</td>
<td>By default, your instance can access the Internet. Your instance receives a public IP address by default. An Internet gateway is attached to your default VPC, and your default subnet has a route to the Internet gateway.</td>
<td>By default, your instance cannot access the Internet. Your instance doesn't receive a public IP address by default. Your VPC may have an Internet gateway, depending on how it was created.</td>
</tr>
<tr>
<td>IPv6 addressing</td>
<td>IPv6 addressing is not supported. You cannot assign IPv6 addresses to your instances.</td>
<td>You can optionally associate an IPv6 CIDR block with your VPC, and assign IPv6 addresses to instances in your VPC.</td>
<td>You can optionally associate an IPv6 CIDR block with your VPC, and assign IPv6 addresses to instances in your VPC.</td>
</tr>
</tbody>
</table>

### Security Groups for EC2-Classic

If you're using EC2-Classic, you must use security groups created specifically for EC2-Classic. When you launch an instance in EC2-Classic, you must specify a security group in the same region as the instance. You can't specify a security group that you created for a VPC when you launch an instance in EC2-Classic.
After you launch an instance in EC2-Classic, you can't change its security groups. However, you can add rules to or remove rules from a security group, and those changes are automatically applied to all instances that are associated with the security group after a short period.

Your AWS account automatically has a default security group per region for EC2-Classic. If you try to delete the default security group, you'll get the following error: Client.InvalidGroup.Reserved: The security group 'default' is reserved.

You can create custom security groups. The security group name must be unique within your account for the region. To create a security group for use in EC2-Classic, choose No VPC for the VPC.

You can add inbound rules to your default and custom security groups. You can't change the outbound rules for an EC2-Classic security group. When you create a security group rule, you can use a different security group for EC2-Classic in the same region as the source or destination. To specify a security group for another AWS account, add the AWS account ID as a prefix; for example, 111122223333/sg-edcd9784.

In EC2-Classic, you can have up to 500 security groups in each region for each account. You can associate an instance with up to 500 security groups and add up to 100 rules to a security group.

**IP Addressing and DNS**

Amazon provides a DNS server that resolves Amazon-provided IPv4 DNS hostnames to IPv4 addresses. In EC2-Classic, the Amazon DNS server is located at 172.16.0.23.

If you create a custom firewall configuration in EC2-Classic, you must create a rule in your firewall that allows inbound traffic from port 53 (DNS)—with a destination port from the ephemeral range—from the address of the Amazon DNS server; otherwise, internal DNS resolution from your instances fails. If your firewall doesn't automatically allow DNS query responses, then you need to allow traffic from the IP address of the Amazon DNS server. To get the IP address of the Amazon DNS server, use the following command from within your instance:

```
ipconfig /all | findstr /c:'DNS Servers'
```

**Elastic IP Addresses**

If your account supports EC2-Classic, there's one pool of Elastic IP addresses for use with the EC2-Classic platform and another for use with your VPCs. You can't associate an Elastic IP address that you allocated for use with a VPC with an instance in EC2-Classic, and vice-versa. However, you can migrate an Elastic IP address you've allocated for use in the EC2-Classic platform for use with a VPC. You cannot migrate an Elastic IP address to another region.

**To allocate an Elastic IP address for use in EC2-Classic using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Choose Allocate new address.
4. Select Classic, and then choose Allocate. Close the confirmation screen.

**Migrating an Elastic IP Address from EC2-Classic**

If your account supports EC2-Classic, you can migrate Elastic IP addresses that you've allocated for use with EC2-Classic platform to be used with a VPC, within the same region. This can assist you to migrate your resources from EC2-Classic to a VPC; for example, you can launch new web servers in your VPC, and
then use the same Elastic IP addresses that you used for your web servers in EC2-Classic for your new VPC web servers.

After you've migrated an Elastic IP address to a VPC, you cannot use it with EC2-Classic. However, if required, you can restore it to EC2-Classic. You cannot migrate an Elastic IP address that was originally allocated for use with a VPC to EC2-Classic.

To migrate an Elastic IP address, it must not be associated with an instance. For more information about disassociating an Elastic IP address from an instance, see Disassociating an Elastic IP Address and Reassociating with a Different Instance (p. 677).

You can migrate as many EC2-Classic Elastic IP addresses as you can have in your account. However, when you migrate an Elastic IP address, it counts against your Elastic IP address limit for VPCs. You cannot migrate an Elastic IP address if it will result in your exceeding your limit. Similarly, when you restore an Elastic IP address to EC2-Classic, it counts against your Elastic IP address limit for EC2-Classic. For more information, see Elastic IP Address Limit (p. 679).

You cannot migrate an Elastic IP address that has been allocated to your account for less than 24 hours.

You can migrate an Elastic IP address from EC2-Classic using the Amazon EC2 console or the Amazon VPC console. This option is only available if your account supports EC2-Classic.

**To move an Elastic IP address using the Amazon EC2 console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Select the Elastic IP address, and choose Actions, Move to VPC scope.
4. In the confirmation dialog box, choose Move Elastic IP.

You can restore an Elastic IP address to EC2-Classic using the Amazon EC2 console or the Amazon VPC console.

**To restore an Elastic IP address to EC2-Classic using the Amazon EC2 console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic IPs.
3. Select the Elastic IP address, choose Actions, Restore to EC2 scope.
4. In the confirmation dialog box, choose Restore.

After you've performed the command to move or restore your Elastic IP address, the process of migrating the Elastic IP address can take a few minutes. Use the describe-moving-addresses command to check whether your Elastic IP address is still moving, or has completed moving.

After you've moved your Elastic IP address, you can view its allocation ID on the Elastic IPs page in the Allocation ID field.

If the Elastic IP address is in a moving state for longer than 5 minutes, contact Premium Support.

**To move an Elastic IP address using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- move-address-to-vpc (AWS CLI)
- Move-EC2AddressToVpc (AWS Tools for Windows PowerShell)
To restore an Elastic IP address to EC2-Classic using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `restore-address-to-classic` (AWS CLI)
- `Restore-EC2AddressToClassic` (AWS Tools for Windows PowerShell)

To describe the status of your moving addresses using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `describe-moving-addresses` (AWS CLI)
- `Get-EC2Address` (AWS Tools for Windows PowerShell)

Sharing and Accessing Resources Between EC2-Classic and a VPC

Some resources and features in your AWS account can be shared or accessed between EC2-Classic and a VPC, for example, through ClassicLink. For more information, see ClassicLink (p. 760).

If your account supports EC2-Classic, you might have set up resources for use in EC2-Classic. If you want to migrate from EC2-Classic to a VPC, you must recreate those resources in your VPC. For more information about migrating from EC2-Classic to a VPC, see Migrating from a Windows Instance in EC2-Classic to a Windows Instance in a VPC (p. 772).

The following resources can be shared or accessed between EC2-Classic and a VPC.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI</td>
<td></td>
</tr>
<tr>
<td>Bundle task</td>
<td></td>
</tr>
<tr>
<td>EBS volume</td>
<td></td>
</tr>
<tr>
<td>Elastic IP address (IPv4)</td>
<td>You can migrate an Elastic IP address from EC2-Classic to a VPC. You can't migrate an Elastic IP address that was originally allocated for use in a VPC to EC2-Classic. For more information, see Migrating an Elastic IP Address from EC2-Classic (p. 757).</td>
</tr>
<tr>
<td>Instance</td>
<td>An EC2-Classic instance can communicate with instances in a VPC using public IPv4 addresses, or you can use ClassicLink to enable communication over private IPv4 addresses. You can't migrate an instance from EC2-Classic to a VPC. However, you can migrate your application from an instance in EC2-Classic to an instance in a VPC. For more information, see Migrating from a Windows Instance in EC2-Classic to a Windows Instance in a VPC (p. 772).</td>
</tr>
<tr>
<td>Resource</td>
<td>Notes</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Key pair</td>
<td>If you're using ClassicLink, you can register a linked EC2-Classic instance with a load balancer in a VPC, provided that the VPC has a subnet in the same Availability Zone as the instance. You can't migrate a load balancer from EC2-Classic to a VPC. You can't register an instance in a VPC with a load balancer in EC2-Classic.</td>
</tr>
<tr>
<td>Load balancer</td>
<td></td>
</tr>
<tr>
<td>Placement group</td>
<td>You can change the network platform for your Reserved Instances from EC2-Classic to a VPC. For more information, see Modifying Reserved Instances (p. 220). You can't migrate a security group from EC2-Classic to a VPC. You can copy rules from a security group for EC2-Classic to a security group for a VPC. For more information, see Creating a Security Group (p. 587).</td>
</tr>
<tr>
<td>Reserved Instance</td>
<td></td>
</tr>
<tr>
<td>Security group</td>
<td>A linked EC2-Classic instance can use a VPC security groups through ClassicLink to control traffic to and from the VPC. VPC instances can't use EC2-Classic security groups. You can't migrate a security group from EC2-Classic to a VPC. You can copy rules from a security group for EC2-Classic to a security group for a VPC. For more information, see Creating a Security Group (p. 587).</td>
</tr>
<tr>
<td>Snapshot</td>
<td></td>
</tr>
</tbody>
</table>

The following resources can't be shared or moved between EC2-Classic and a VPC:

- Spot Instances

---

**ClassicLink**

ClassicLink allows you to link EC2-Classic instances to a VPC in your account, within the same region. If you associate the VPC security groups with a EC2-Classic instance, this enables communication between your EC2-Classic instance and instances in your VPC using private IPv4 addresses. ClassicLink removes the need to make use of public IPv4 addresses or Elastic IP addresses to enable communication between instances in these platforms.

ClassicLink is available to all users with accounts that support the EC2-Classic platform, and can be used with any EC2-Classic instance. For more information about migrating your resources to a VPC, see Migrating from a Windows Instance in EC2-Classic to a Windows Instance in a VPC (p. 772).

There is no additional charge for using ClassicLink. Standard charges for data transfer and instance usage apply.

**Contents**

- ClassicLink Basics (p. 761)
- ClassicLink Limitations (p. 763)
- Working with ClassicLink (p. 764)
ClassicLink Basics

There are two steps to linking an EC2-Classic instance to a VPC using ClassicLink. First, you must enable the VPC for ClassicLink. By default, all VPCs in your account are not enabled for ClassicLink, to maintain their isolation. After you've enabled the VPC for ClassicLink, you can then link any running EC2-Classic instance in the same region in your account to that VPC. Linking your instance includes selecting security groups from the VPC to associate with your EC2-Classic instance. After you've linked the instance, it can communicate with instances in your VPC using their private IP addresses, provided the VPC security groups allow it. Your EC2-Classic instance does not lose its private IP address when linked to the VPC.

**Note**

Linking your instance to a VPC is sometimes referred to as *attaching* your instance.

A linked EC2-Classic instance can communicate with instances in a VPC, but it does not form part of the VPC. If you list your instances and filter by VPC, for example, through the `DescribeInstances` API request, or by using the **Instances** screen in the Amazon EC2 console, the results do not return any EC2-Classic instances that are linked to the VPC. For more information about viewing your linked EC2-Classic instances, see Viewing Your ClassicLink-Enabled VPCs and Linked Instances (p. 765).

By default, if you use a public DNS hostname to address an instance in a VPC from a linked EC2-Classic instance, the hostname resolves to the instance's public IP address. The same occurs if you use a public DNS hostname to address a linked EC2-Classic instance from an instance in the VPC. If you want the public DNS hostname to resolve to the private IP address, you can enable ClassicLink DNS support for the VPC. For more information, see Enabling ClassicLink DNS Support (p. 766).

If you no longer require a ClassicLink connection between your instance and the VPC, you can unlink the EC2-Classic instance from the VPC. This disassociates the VPC security groups from the EC2-Classic instance. A linked EC2-Classic instance is automatically unlinked from a VPC when it's stopped. After you've unlinked all linked EC2-Classic instances from the VPC, you can disable ClassicLink for the VPC.

Using Other AWS Services in Your VPC With ClassicLink

Linked EC2-Classic instances can access the following AWS services in the VPC: Amazon Redshift, Amazon ElastiCache, Elastic Load Balancing, and Amazon RDS. However, instances in the VPC cannot access the AWS services provisioned by the EC2-Classic platform using ClassicLink.

If you use Elastic Load Balancing, you can register your linked EC2-Classic instances with the load balancer. You must create your load balancer in the ClassicLink-enabled VPC and enable the Availability Zone in which the instance runs. If you terminate the linked EC2-Classic instance, the load balancer deregisters the instance.

If you use Amazon EC2 Auto Scaling, you can create an Amazon EC2 Auto Scaling group with instances that are automatically linked to a specified ClassicLink-enabled VPC at launch. For more information, see Linking EC2-Classic Instances to a VPC in the Amazon EC2 Auto Scaling User Guide.

If you use Amazon RDS instances or Amazon Redshift clusters in your VPC, and they are publicly accessible (accessible from the Internet), the endpoint you use to address those resources from a linked EC2-Classic instance by default resolves to a public IP address. If those resources are not publicly accessible, the endpoint resolves to a private IP address. To address a publicly accessible RDS instance or Redshift cluster over private IP using ClassicLink, you must use their private IP address or private DNS hostname, or you must enable ClassicLink DNS support for the VPC.

If you use a private DNS hostname or a private IP address to address an RDS instance, the linked EC2-Classic instance cannot use the failover support available for Multi-AZ deployments.
You can use the Amazon EC2 console to find the private IP addresses of your Amazon Redshift, Amazon ElastiCache, or Amazon RDS resources.

**To locate the private IP addresses of AWS resources in your VPC**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Network Interfaces**.
3. Check the descriptions of the network interfaces in the **Description** column. A network interface that's used by Amazon Redshift, Amazon ElastiCache, or Amazon RDS will have the name of the service in the description. For example, a network interface that's attached to an Amazon RDS instance will have the following description: RDSNetworkInterface.
4. Select the required network interface.
5. In the details pane, get the private IP address from the **Primary private IPv4 IP** field.

**Controlling the Use of ClassicLink**

By default, IAM users do not have permission to work with ClassicLink. You can create an IAM policy that grants users permissions to enable or disable a VPC for ClassicLink, link or unlink an instance to a ClassicLink-enabled VPC, and to view ClassicLink-enabled VPCs and linked EC2-Classic instances. For more information about IAM policies for Amazon EC2, see IAM Policies for Amazon EC2 (p. 599).

For more information about policies for working with ClassicLink, see the following example: Example IAM Policies for ClassicLink (p. 767).

**Security Groups in ClassicLink**

Linking your EC2-Classic instance to a VPC does not affect your EC2-Classic security groups. They continue to control all traffic to and from the instance. This excludes traffic to and from instances in the VPC, which is controlled by the VPC security groups that you associated with the EC2-Classic instance. EC2-Classic instances that are linked to the same VPC cannot communicate with each other through the VPC; regardless of whether they are associated with the same VPC security group. Communication between EC2-Classic instances is controlled by the EC2-Classic security groups associated with those instances. For an example of a security group configuration, see Example: ClassicLink Security Group Configuration for a Three-Tier Web Application (p. 769).

After you've linked your instance to a VPC, you cannot change which VPC security groups are associated with the instance. To associate different security groups with your instance, you must first unlink the instance, and then link it to the VPC again, choosing the required security groups.

**Routing for ClassicLink**

When you enable a VPC for ClassicLink, a static route is added to all of the VPC route tables with a destination of 10.0.0.0/8 and a target of local. This allows communication between instances in the VPC and any EC2-Classic instances that are then linked to the VPC. If you add a custom route table to a ClassicLink-enabled VPC, a static route is automatically added with a destination of 10.0.0.0/8 and a target of local. When you disable ClassicLink for a VPC, this route is automatically deleted in all of the VPC route tables.

VPCs that are in the 10.0.0.0/16 and 10.1.0.0/16 IP address ranges can be enabled for ClassicLink only if they do not have any existing static routes in route tables in the 10.0.0.0/8 IP address range, excluding the local routes that were automatically added when the VPC was created. Similarly, if you've enabled a VPC for ClassicLink, you may not be able to add any more specific routes to your route tables within the 10.0.0.0/8 IP address range.

**Important**

If your VPC CIDR block is a publicly routable IP address range, consider the security implications before you link an EC2-Classic instance to your VPC. For example, if your linked EC2-Classic instance receives an incoming Denial of Service (DoS) request flood attack from a source IP
address that falls within the VPC's IP address range, the response traffic is sent into your VPC.
We strongly recommend that you create your VPC using a private IP address range as specified
in RFC 1918.

For more information about route tables and routing in your VPC, see Route Tables in the Amazon VPC
User Guide.

Enabling a VPC Peering Connection for ClassicLink

If you have a VPC peering connection between two VPCs, and there are one or more EC2-Classic
instances that are linked to one or both of the VPCs via ClassicLink, you can extend the VPC peering
connection to enable communication between the EC2-Classic instances and the instances in the
VPC on the other side of the VPC peering connection. This enables the EC2-Classic instances and the
instances in the VPC to communicate using private IP addresses. To do this, you can enable a local VPC
to communicate with a linked EC2-Classic instance in a peer VPC, or you can enable a local linked EC2-
Classic instance to communicate with instances in a peer VPC.

If you enable a local VPC to communicate with a linked EC2-Classic instance in a peer VPC, a static route
is automatically added to your route tables with a destination of 10.0.0.0/8 and a target of local.

For more information and examples, see Configurations With ClassicLink in the Amazon VPC Peering
Guide.

ClassicLink Limitations

To use the ClassicLink feature, you need to be aware of the following limitations:

• You can link an EC2-Classic instance to only one VPC at a time.
• If you stop your linked EC2-Classic instance, it's automatically unlinked from the VPC and the VPC
  security groups are no longer associated with the instance. You can link your instance to the VPC again
  after you've restarted it.
• You cannot link an EC2-Classic instance to a VPC that's in a different region or a different AWS account.
• You cannot use ClassicLink to link a VPC instance to a different VPC, or to a EC2-Classic resource.
  To establish a private connection between VPCs, you can use a VPC peering connection. For more
  information, see the Amazon VPC Peering Guide.
• You cannot associate a VPC Elastic IP address with a linked EC2-Classic instance.
• You cannot enable EC2-Classic instances for IPv6 communication. You can associate an IPv6 CIDR block
  with your VPC and assign IPv6 address to resources in your VPC, however, communication between a
  ClassicLinked instance and resources in the VPC is over IPv4 only.
• VPCs with routes that conflict with the EC2-Classic private IP address range of 10/8 cannot be
  enabled for ClassicLink. This does not include VPCs with 10.0.0.0/16 and 10.1.0.0/16 IP address
  ranges that already have local routes in their route tables. For more information, see Routing for
  ClassicLink (p. 762).
• VPCs configured for dedicated hardware tenancy cannot be enabled for ClassicLink. Contact AWS
  support to request that your dedicated tenancy VPC be allowed to be enabled for ClassicLink.

  Important
  EC2-Classic instances are run on shared hardware. If you've set the tenancy of your VPC to
dedicated because of regulatory or security requirements, then linking an EC2-Classic
instance to your VPC might not conform to those requirements, as this allows a shared
 tenancy resource to address your isolated resources directly using private IP addresses. If you
 need to enable your dedicated VPC for ClassicLink, provide a detailed reason in your request
to AWS support.
• If you link your EC2-Classic instance to a VPC in the 172.16.0.0/16 range, and you have a DNS
  server running on the 172.16.0.23/32 IP address within the VPC, then your linked EC2-Classic
  instance can't access the VPC DNS server. To work around this issue, run your DNS server on a different
  IP address within the VPC.
• ClassicLink doesn't support transitive relationships out of the VPC. Your linked EC2-Classic instance doesn't have access to any VPN connection, VPC gateway endpoint, NAT gateway, or Internet gateway associated with the VPC. Similarly, resources on the other side of a VPN connection or an Internet gateway don't have access to a linked EC2-Classic instance.

Working with ClassicLink

You can use the Amazon EC2 and Amazon VPC consoles to work with the ClassicLink feature. You can enable or disable a VPC for ClassicLink, and link and unlink EC2-Classic instances to a VPC.

Note
The ClassicLink features are only visible in the consoles for accounts and regions that support EC2-Classic.

Tasks
• Enabling a VPC for ClassicLink (p. 764)
• Creating a VPC with ClassicLink Enabled (p. 764)
• Linking an Instance to a VPC (p. 765)
• Linking an Instance to a VPC at Launch (p. 765)
• Viewing Your ClassicLink-Enabled VPCs and Linked Instances (p. 765)
• Enabling ClassicLink DNS Support (p. 766)
• Disabling ClassicLink DNS Support (p. 766)
• Unlinking an Instance from a VPC (p. 767)
• Disabling ClassicLink for a VPC (p. 767)

Enabling a VPC for ClassicLink

To link an EC2-Classic instance to a VPC, you must first enable the VPC for ClassicLink. You cannot enable a VPC for ClassicLink if the VPC has routing that conflicts with the EC2-Classic private IP address range. For more information, see Routing for ClassicLink (p. 762).

To enable a VPC for ClassicLink

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose Your VPCs.
3. Choose a VPC, and then choose Actions, Enable ClassicLink.
4. In the confirmation dialog box, choose Yes, Enable.
5. (Optional) If you want the public DNS hostname to resolve to the private IP address, enable ClassicLink DNS support for the VPC before you link any instances. For more information, see Enabling ClassicLink DNS Support (p. 766).

Creating a VPC with ClassicLink Enabled

You can create a new VPC and immediately enable it for ClassicLink by using the VPC wizard in the Amazon VPC console.

To create a VPC with ClassicLink enabled

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. From the Amazon VPC dashboard, choose Start VPC Wizard.
3. Select one of the VPC configuration options and choose Select.
4. On the next page of the wizard, choose Yes for Enable ClassicLink. Complete the rest of the steps in the wizard to create your VPC. For more information about using the VPC wizard, see Scenarios for Amazon VPC in the Amazon VPC User Guide.

5. (Optional) If you want the public DNS hostname to resolve to the private IP address, enable ClassicLink DNS support for the VPC before you link any instances. For more information, see Enabling ClassicLink DNS Support (p. 766).

### Linking an Instance to a VPC

After you’ve enabled a VPC for ClassicLink, you can link an EC2-Classic instance to it.

**Note**

You can only link a running EC2-Classic instance to a VPC. You cannot link an instance that’s in the stopped state.

If you want the public DNS hostname to resolve to the private IP address, enable ClassicLink DNS support for the VPC before you link the instance. For more information, see Enabling ClassicLink DNS Support (p. 766).

**To link an instance to a VPC**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the running EC2-Classic instance, choose Actions, ClassicLink, Link to VPC. You can select more than one instance to link to the same VPC.
4. In the dialog box that displays, select a VPC from the list. Only VPCs that have been enabled for ClassicLink are displayed.
5. Select one or more of the VPC security groups to associate with your instance. When you are done, choose Link to VPC.

### Linking an Instance to a VPC at Launch

You can use the launch wizard in the Amazon EC2 console to launch an EC2-Classic instance and immediately link it to a ClassicLink-enabled VPC.

**To link an instance to a VPC at launch**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the Amazon EC2 dashboard, choose Launch Instance.
3. Select an AMI, and then choose an instance type. On the Configure Instance Details page, ensure that you select Launch into EC2-Classic from the Network list.
   
   **Note**
   
   Some instance types, such as T2 instance types, can only be launched into a VPC. Ensure that you select an instance type that can be launched into EC2-Classic.
4. In the Link to VPC (ClassicLink) section, select a VPC from Link to VPC. Only ClassicLink-enabled VPCs are displayed. Select the security groups from the VPC to associate with the instance. Complete the other configuration options on the page, and then complete the rest of the steps in the wizard to launch your instance. For more information about using the launch wizard, see Launching Your Instance from an AMI (p. 334).

### Viewing Your ClassicLink-Enabled VPCs and Linked Instances

You can view all of your ClassicLink-enabled VPCs in the Amazon VPC console, and your linked EC2-Classic instances in the Amazon EC2 console.
To view your ClassicLink-enabled VPCs

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose Your VPCs.
3. Select a VPC, and in the Summary tab, look for the ClassicLink field. A value of Enabled indicates that the VPC is enabled for ClassicLink.
4. Alternatively, look for the ClassicLink column, and view the value that's displayed for each VPC (Enabled or Disabled). If the column is not visible, choose Edit Table Columns (the gear-shaped icon), select the ClassicLink attribute, and then choose Close.

To view your linked EC2-Classic instances

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select an EC2-Classic instance, and in the Description tab, look for the ClassicLink field. If the instance is linked to a VPC, the field displays the ID of the VPC to which the instance is linked. If the instance is not linked to any VPC, the field displays Unlinked.
4. Alternatively, you can filter your instances to display only linked EC2-Classic instances for a specific VPC or security group. In the search bar, start typing ClassicLink, select the relevant ClassicLink resource attribute, and then select the security group ID or the VPC ID.

Enabling ClassicLink DNS Support

You can enable ClassicLink DNS support for your VPC so that DNS hostnames that are addressed between linked EC2-Classic instances and instances in the VPC resolve to private IP addresses and not public IP addresses. For this feature to work, your VPC must be enabled for DNS hostnames and DNS resolution.

**Note**
If you enable ClassicLink DNS support for your VPC, your linked EC2-Classic instance can access any private hosted zone associated with the VPC. For more information, see Working with Private Hosted Zones in the Amazon Route 53 Developer Guide.

To enable ClassicLink DNS support

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose Your VPCs.
3. Select your VPC, and choose Actions, Edit ClassicLink DNS Support.
4. Choose Yes to enable ClassicLink DNS support, and choose Save.

Disabling ClassicLink DNS Support

You can disable ClassicLink DNS support for your VPC so that DNS hostnames that are addressed between linked EC2-Classic instances and instances in the VPC resolve to public IP addresses and not private IP addresses.

To disable ClassicLink DNS support

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose Your VPCs.
3. Select your VPC, and choose Actions, Edit ClassicLink DNS Support.
4. Choose No to disable ClassicLink DNS support, and choose Save.
Unlinking an Instance from a VPC

If you no longer require a ClassicLink connection between your EC2-Classic instance and your VPC, you can unlink the instance from the VPC. Unlinking the instance disassociates the VPC security groups from the instance.

**Note**

A stopped instance is automatically unlinked from a VPC.

**To unlink an instance from a VPC**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**, and select your instance.
3. In the **Actions** list, select **ClassicLink, Unlink Instance**. You can select more than one instance to unlink from the same VPC.
4. Choose **Yes** in the confirmation dialog box.

Disabling ClassicLink for a VPC

If you no longer require a connection between EC2-Classic instances and your VPC, you can disable ClassicLink on the VPC. You must first unlink all linked EC2-Classic instances that are linked to the VPC.

**To disable ClassicLink for a VPC**

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose **Your VPCs**.
3. Select your VPC, then choose **Actions, Disable ClassicLink**.
4. In the confirmation dialog box, choose **Yes, Disable**.

Example IAM Policies for ClassicLink

You can enable a VPC for ClassicLink and then link an EC2-Classic instance to the VPC. You can also view your ClassicLink-enabled VPCs, and all of your EC2-Classic instances that are linked to a VPC. You can create policies with resource-level permission for the `ec2:EnableVpcClassicLink`, `ec2:DisableVpcClassicLink`, `ec2:AttachClassicLinkVpc`, and `ec2:DetachClassicLinkVpc` actions to control how users are able to use those actions. Resource-level permissions are not supported for `ec2:Describe*` actions.

**Examples**

- Full Permissions to Work with ClassicLink (p. 767)
- Enable and Disable a VPC for ClassicLink (p. 768)
- Link Instances (p. 768)
- Unlink Instances (p. 769)

**Full Permissions to Work with ClassicLink**

The following policy grants users permissions to view ClassicLink-enabled VPCs and linked EC2-Classic instances, to enable and disable a VPC for ClassicLink, and to link and unlink instances from a ClassicLink-enabled VPC.

```
Enable and Disable a VPC for ClassicLink

The following policy allows user to enable and disable VPCs for ClassicLink that have the specific tag 'purpose=classiclink'. Users cannot enable or disable any other VPCs for ClassicLink.

```
{  
  "Version": "2012-10-17",
  "Statement": [  
    {  
      "Effect": "Allow",
      "Resource": "*"
    }  
  ]
}
```

Link Instances

The following policy grants users permissions to link instances to a VPC only if the instance is an m3.large instance type. The second statement allows users to use the VPC and security group resources, which are required to link an instance to a VPC.

```
{  
  "Version": "2012-10-17",
  "Statement": [  
    {  
      "Effect": "Allow",
      "Action": "ec2:AttachClassicLinkVpc",
      "Condition": {  
        "StringEquals": {  
          "ec2:ResourceTag/purpose": "classiclink"
        }
      }
    },  
    {  
      "Effect": "Allow",
      "Action": "ec2:AttachClassicLinkVpc",
      "Resource": [  
        "arn:aws:ec2:region:account:vpc/*",
        "arn:aws:ec2:region:account:security-group/"
      ]
    }  
  ]
}
```
The following policy grants users permissions to link instances to a specific VPC (vpc-1a2b3c4d) only, and to associate only specific security groups from the VPC to the instance (sg-1122aabb and sg-aabb2233). Users cannot link an instance to any other VPC, and they cannot specify any other of the VPC security groups to associate with the instance in the request.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "ec2:AttachClassicLinkVpc",
            "Resource": [
                "arn:aws:ec2:region:account:vpc/vpc-1a2b3c4d",
                "arn:aws:ec2:region:account:instance/*",
                "arn:aws:ec2:region:account:security-group/sg-aabb2233"
            ]
        }
    ]
}
```

### Unlink Instances

The following grants users permission to unlink any linked EC2-Classic instance from a VPC, but only if the instance has the tag "unlink=true". The second statement grants users permissions to use the VPC resource, which is required to unlink an instance from a VPC.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "ec2:DetachClassicLinkVpc",
            "Resource": [
                "arn:aws:ec2:region:account:instance/*"
            ],
            "Condition": {
                "StringEquals": {
                    "ec2:ResourceTag/unlink":"true"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "ec2:DetachClassicLinkVpc",
            "Resource": [
                "arn:aws:ec2:region:account:vpc/*"
            ]
        }
    ]
}
```

### Example: ClassicLink Security Group Configuration for a Three-Tier Web Application

In this example, you have an application with three instances: a public-facing web server, an application server, and a database server. Your web server accepts HTTPS traffic from the Internet, and then communicates with your application server over TCP port 6001. Your application server then
communicates with your database server over TCP port 6004. You're in the process of migrating your entire application to a VPC in your account. You've already migrated your application server and your database server to your VPC. Your web server is still in EC2-Classic and linked to your VPC via ClassicLink.

You want a security group configuration that allows traffic to flow only between these instances. You have four security groups: two for your web server (sg-1a1a1a1a and sg-2b2b2b2b), one for your application server (sg-3c3c3c3c), and one for your database server (sg-4d4d4d4d).

The following diagram displays the architecture of your instances, and their security group configuration.

Security Groups for Your Web Server (sg-1a1a1a1a and sg-2b2b2b2b)

You have one security group in EC2-Classic, and the other in your VPC. You associated the VPC security group with your web server instance when you linked the instance to your VPC via ClassicLink. The VPC security group enables you to control the outbound traffic from your web server to your application server.

The following are the security group rules for the EC2-Classic security group (sg-1a1a1a1a).

<table>
<thead>
<tr>
<th>Inbound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
### Amazon Elastic Compute Cloud
#### User Guide for Windows Instances
**ClassicLink**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Port Range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.0.0/0</td>
<td>HTTPS</td>
<td>443</td>
<td>Allows Internet traffic to reach your web server.</td>
</tr>
</tbody>
</table>

The following are the security group rules for the VPC security group (`sg-2b2b2b2b`).

<table>
<thead>
<tr>
<th>Outbound</th>
<th>Destination</th>
<th>Type</th>
<th>Port Range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><code>sg-3c3c3c3c</code></td>
<td>TCP</td>
<td>6001</td>
<td>Allows outbound traffic from your web server to your application server in your VPC (or to any other instance associated with <code>sg-3c3c3c3c</code>).</td>
</tr>
</tbody>
</table>

**Security Group for Your Application Server (`sg-3c3c3c3c`)**

The following are the security group rules for the VPC security group that's associated with your application server.

<table>
<thead>
<tr>
<th>Inbound</th>
<th>Source</th>
<th>Type</th>
<th>Port Range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><code>sg-2b2b2b2b</code></td>
<td>TCP</td>
<td>6001</td>
<td>Allows the specified type of traffic from your web server (or any other instance associated with <code>sg-2b2b2b2b</code>) to reach your application server.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outbound</th>
<th>Destination</th>
<th>Type</th>
<th>Port Range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><code>sg-4d4d4d4d</code></td>
<td>TCP</td>
<td>6004</td>
<td>Allows outbound traffic from the application server to the database server (or to any other instance associated with <code>sg-4d4d4d4d</code>).</td>
</tr>
</tbody>
</table>

**Security Group for Your Database Server (`sg-4d4d4d4d`)**

The following are the security group rules for the VPC security group that's associated with your database server.

<table>
<thead>
<tr>
<th>Inbound</th>
<th>Source</th>
<th>Type</th>
<th>Port Range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><code>sg-3c3c3c3c</code></td>
<td>TCP</td>
<td>6004</td>
<td>Allows the specified type of traffic from your application server (or any other instance associated with <code>sg-3c3c3c3c</code>) to reach your database server.</td>
</tr>
</tbody>
</table>
Migrating from a Windows Instance in EC2-Classic to a Windows Instance in a VPC

If you created your AWS account before 2013-12-04, you might have support for EC2-Classic in some regions. Some Amazon EC2 resources and features, such as enhanced networking and newer instance types, require a virtual private cloud (VPC). Some resources can be shared between EC2-Classic and a VPC, while some can't. For more information, see Sharing and Accessing Resources Between EC2-Classic and a VPC (p. 759).

If your account supports EC2-Classic, you might have set up resources for use in EC2-Classic. If you want to migrate from EC2-Classic to a VPC, you must recreate those resources in your VPC.

There are two ways of migrating to a VPC. You can do a full migration, or you can do an incremental migration over time. The method you choose depends on the size and complexity of your application in EC2-Classic. For example, if your application consists of one or two instances running a static website, and you can afford a short period of downtime, you can do a full migration. If you have a multi-tier application with processes that cannot be interrupted, you can do an incremental migration using ClassicLink. This allows you to transfer functionality one component at a time until your application is running fully in your VPC.

If you need to migrate a Linux instance, see Migrating a Linux Instance from EC2-Classic to a VPC in the Amazon EC2 User Guide for Linux Instances.

Contents

- Full Migration to a VPC (p. 772)
- Incremental Migration to a VPC Using ClassicLink (p. 777)

Full Migration to a VPC

Complete the following tasks to fully migrate your application from EC2-Classic to a VPC.

Tasks

- Step 1: Create a VPC (p. 772)
- Step 2: Configure Your Security Group (p. 773)
- Step 3: Create an AMI from Your EC2-Classic Instance (p. 773)
- Step 4: Launch an Instance Into Your VPC (p. 774)
- Example: Migrating a Simple Web Application (p. 775)

Step 1: Create a VPC

To start using a VPC, ensure that you have one in your account. You can create one using one of these methods:

- Your AWS account comes with a default VPC in each region, which is ready for you to use. Instances that you launch are by default launched into this VPC, unless you specify otherwise. For more information about your default VPC, see Your Default VPC and Subnets. Use this option if you'd prefer not to set up a VPC yourself, or if you do not need specific requirements for your VPC configuration.
- In your existing AWS account, open the Amazon VPC console and use the VPC wizard to create a new VPC. For more information, see Scenarios for Amazon VPC. Use this option if you want to set up a VPC quickly in your existing EC2-Classic account, using one of the available configuration sets in the wizard. You'll specify this VPC each time you launch an instance.
- In your existing AWS account, open the Amazon VPC console and set up the components of a VPC according to your requirements. For more information, see Your VPC and Subnets. Use this option if
you have specific requirements for your VPC, such as a particular number of subnets. You'll specify this VPC each time you launch an instance.

**Step 2: Configure Your Security Group**

You cannot use the same security groups between EC2-Classic and a VPC. However, if you want your instances in your VPC to have the same security group rules as your EC2-Classic instances, you can use the Amazon EC2 console to copy your existing EC2-Classic security group rules to a new VPC security group.

**Important**

You can only copy security group rules to a new security group in the same AWS account in the same region. If you've created a new AWS account, you cannot use this method to copy your existing security group rules to your new account. You'll have to create a new security group, and add the rules yourself. For more information about creating a new security group, see Amazon EC2 Security Groups for Windows Instances (p. 583).

**To copy your security group rules to a new security group**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Security Groups**.
3. Select the security group that's associated with your EC2-Classic instance, then choose **Actions** and select **Copy to new**.
4. In the **Create Security Group** dialog box, specify a name and description for your new security group. Select your VPC from the **VPC** list.
5. The **Inbound** tab is populated with the rules from your EC2-Classic security group. You can modify the rules as required. In the **Outbound** tab, a rule that allows all outbound traffic has automatically been created for you. For more information about modifying security group rules, see Amazon EC2 Security Groups for Windows Instances (p. 583).

**Note**

If you've defined a rule in your EC2-Classic security group that references another security group, you will not be able to use the same rule in your VPC security group. Modify the rule to reference a security group in the same VPC.

6. Choose **Create**.

**Step 3: Create an AMI from Your EC2-Classic Instance**

An AMI is a template for launching your instance. You can create your own AMI based on an existing EC2-Classic instance, then use that AMI to launch instances into your VPC.

For more information, see Creating a Custom Windows AMI (p. 64).

**(Optional) Store Your Data on Amazon EBS Volumes**

You can create an Amazon EBS volume and use it to back up and store the data on your instance—like you would use a physical hard drive. Amazon EBS volumes can be attached and detached from any instance in the same Availability Zone. You can detach a volume from your instance in EC2-Classic, and attach it to a new instance that you launch into your VPC in the same Availability Zone.

For more information about Amazon EBS volumes, see the following topics:

- Amazon EBS Volumes (p. 783)
- Creating an Amazon EBS Volume (p. 800)
- Attaching an Amazon EBS Volume to an Instance (p. 803)
To back up the data on your Amazon EBS volume, you can take periodic snapshots of your volume. If you need to, you can restore an Amazon EBS volume from your snapshot. For more information about Amazon EBS snapshots, see the following topics:

- Amazon EBS Snapshots (p. 818)
- Creating Amazon EBS Snapshots (p. 822)
- Restoring an Amazon EBS Volume from a Snapshot (p. 802)

**Step 4: Launch an Instance Into Your VPC**

After you've created an AMI, you can launch an instance into your VPC. The instance will have the same data and configurations as your existing EC2-Classic instance.

You can either launch your instance into a VPC that you've created in your existing account, or into a new, VPC-only AWS account.

**Using Your Existing EC2-Classic Account**

You can use the Amazon EC2 launch wizard to launch an instance into your VPC.

**To launch an instance into your VPC**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the dashboard, choose Launch Instance.
3. On the Choose an Amazon Machine Image page, select the My AMIs category, and select the AMI you created.
4. On the Choose an Instance Type page, select the type of instance, and choose Next: Configure Instance Details.
5. On the Configure Instance Details page, select your VPC from the Network list. Select the required subnet from the Subnet list. Configure any other details you require, then go through the next pages of the wizard until you reach the Configure Security Group page.
6. Select Select an existing group, and select the security group you created earlier. Choose Review and Launch.
7. Review your instance details, then choose Launch to specify a key pair and launch your instance.

For more information about the parameters you can configure in each step of the wizard, see Launching an Instance Using the Launch Instance Wizard (p. 334).

**Using Your New, VPC-Only Account**

To launch an instance in your new AWS account, you'll first have to share the AMI you created with your new account. You can then use the Amazon EC2 launch wizard to launch an instance into your default VPC.

**To share an AMI with your new AWS account**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Switch to the account in which you created your AMI.
3. In the navigation pane, choose AMIs.
4. In the Filter list, ensure Owned by me is selected, then select your AMI.
5. In the Permissions tab, choose Edit. Enter the account number of your new AWS account, choose Add Permission, and then choose Save.
To launch an instance into your default VPC

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Switch to your new AWS account.
3. In the navigation pane, choose AMIs.
4. In the Filter list, select Private images. Select the AMI that you shared from your EC2-Classic account, then choose Launch.
5. On the Choose an Instance Type page, select the type of instance, and choose Next: Configure Instance Details.
6. On the Configure Instance Details page, your default VPC should be selected in the Network list. Configure any other details you require, then go through the next pages of the wizard until you reach the Configure Security Group page.
7. Select Select an existing group, and select the security group you created earlier. Choose Review and Launch.
8. Review your instance details, then choose Launch to specify a key pair and launch your instance.

For more information about the parameters you can configure in each step of the wizard, see Launching an Instance Using the Launch Instance Wizard (p. 334).

Example: Migrating a Simple Web Application

In this example, you use AWS to host your gardening website. To manage your website, you have three running instances in EC2-Classic. Instances A and B host your public-facing web application, and you use Elastic Load Balancing to load balance the traffic between these instances. You've assigned Elastic IP addresses to instances A and B so that you have static IP addresses for configuration and administration tasks on those instances. Instance C holds your MySQL database for your website. You've registered the domain name www.garden.example.com, and you've used Route 53 to create a hosted zone with an alias record set that's associated with the DNS name of your load balancer.

The first part of migrating to a VPC is deciding what kind of VPC architecture will suit your needs. In this case, you've decided on the following: one public subnet for your web servers, and one private subnet for your database server. As your website grows, you can add more web servers and database servers to your subnets. By default, instances in the private subnet cannot access the Internet; however, you can...
enable Internet access through a Network Address Translation (NAT) device in the public subnet. You may want to set up a NAT device to support periodic updates and patches from the Internet for your database server. You'll migrate your Elastic IP addresses to a VPC, and create a load balancer in your public subnet to load balance the traffic between your web servers.

To migrate your web application to a VPC, you can follow these steps:

- **Create a VPC**: In this case, you can use the VPC wizard in the Amazon VPC console to create your VPC and subnets. The second wizard configuration creates a VPC with one private and one public subnet, and launches and configures a NAT device in your public subnet for you. For more information, see Scenario 2: VPC with Public and Private Subnets in the Amazon VPC User Guide.

- **Create AMIs from your instances**: Create an AMI from one of your web servers, and a second AMI from your database server. For more information, see Step 3: Create an AMI from Your EC2-Classic Instance (p. 773).

- **Configure your security groups**: In your EC2-Classic environment, you have one security group for your web servers, and another security group for your database server. You can use the Amazon EC2 console to copy the rules from each security group into new security groups for your VPC. For more information, see Step 2: Configure Your Security Group (p. 773).

  **Tip**
  Create the security groups that are referenced by other security groups first.

- **Launch an instance into your new VPC**: Launch replacement web servers into your public subnet, and launch your replacement database server into your private subnet. For more information, see Step 4: Launch an Instance Into Your VPC (p. 774).
• **Configure your NAT device**: If you are using a NAT instance, you must create security group for it that allows HTTP and HTTPS traffic from your private subnet. For more information, see [NAT Instances](#). If you are using a NAT gateway, traffic from your private subnet is automatically allowed.

• **Configure your database**: When you created an AMI from your database server in EC2-Classic, all the configuration information that was stored in that instance was copied to the AMI. You may have to connect to your new database server and update the configuration details; for example, if you configured your database to grant full read, write, and modification permissions to your web servers in EC2-Classic, you'll have to update the configuration files to grant the same permissions to your new VPC web servers instead.

• **Configure your web servers**: Your web servers will have the same configuration settings as your instances in EC2-Classic. For example, if you configured your web servers to use the database in EC2-Classic, update your web servers' configuration settings to point to your new database instance.

  **Note**
  By default, instances launched into a nondefault subnet are not assigned a public IP address, unless you specify otherwise at launch. Your new database server may not have a public IP address. In this case, you can update your web servers' configuration file to use your new database server's private DNS name. Instances in the same VPC can communicate with each other via private IP address.

• **Migrate your Elastic IP addresses**: Disassociate your Elastic IP addresses from your web servers in EC2-Classic, and then migrate them to a VPC. After you've migrated them, you can associate them with your new web servers in your VPC. For more information, see [Migrating an Elastic IP Address from EC2-Classic](#).

• **Create a new load balancer**: To continue using Elastic Load Balancing to load balance the traffic to your instances, make sure you understand the various ways you can configure your load balancer in VPC. For more information, see [Elastic Load Balancing in Amazon VPC](#).

• **Update your DNS records**: After you've set up your load balancer in your public subnet, ensure that your `www.garden.example.com` domain points to your new load balancer. To do this, you'll need to update your DNS records and update your alias record set in Route 53. For more information about using Route 53, see [Getting Started with Route 53](#).

• **Shut down your EC2-Classic resources**: After you've verified that your web application is working from within the VPC architecture, you can shut down your EC2-Classic resources to stop incurring charges for them. Terminate your EC2-Classic instances, and release your EC2-Classic Elastic IP addresses.

### Incremental Migration to a VPC Using ClassicLink

The ClassicLink feature makes it easier to manage an incremental migration to a VPC. ClassicLink allows you to link an EC2-Classic instance to a VPC in your account in the same region, allowing your new VPC resources to communicate with the EC2-Classic instance using private IPv4 addresses. You can then migrate functionality to the VPC one step at a time. This topic provides some basic steps for managing an incremental migration from EC2-Classic to a VPC.

For more information about ClassicLink, see [ClassicLink](#).

**Topics**

- **Step 1**: Prepare Your Migration Sequence (p. 778)
- **Step 2**: Create a VPC (p. 778)
- **Step 3**: Enable Your VPC for ClassicLink (p. 778)
- **Step 4**: Create an AMI from Your EC2-Classic Instance (p. 778)
- **Step 5**: Launch an Instance Into Your VPC (p. 779)
- **Step 6**: Link Your EC2-Classic Instances to Your VPC (p. 779)
- **Step 7**: Complete the VPC Migration (p. 780)
Step 1: Prepare Your Migration Sequence

To use ClassicLink effectively, you must first identify the components of your application that must be migrated to the VPC, and then confirm the order in which to migrate that functionality.

For example, you have an application that relies on a presentation web server, a backend database server, and authentication logic for transactions. You may decide to start the migration process with the authentication logic, then the database server, and finally, the web server.

Step 2: Create a VPC

To start using a VPC, ensure that you have one in your account. You can create one using one of these methods:

- In your existing AWS account, open the Amazon VPC console and use the VPC wizard to create a new VPC. For more information, see Scenarios for Amazon VPC. Use this option if you want to set up a VPC quickly in your existing EC2-Classic account, using one of the available configuration sets in the wizard. You'll specify this VPC each time you launch an instance.
- In your existing AWS account, open the Amazon VPC console and set up the components of a VPC according to your requirements. For more information, see Your VPC and Subnets. Use this option if you have specific requirements for your VPC, such as a particular number of subnets. You'll specify this VPC each time you launch an instance.

Step 3: Enable Your VPC for ClassicLink

After you've created a VPC, you can enable it for ClassicLink. For more information about ClassicLink, see ClassicLink (p. 760).

To enable a VPC for ClassicLink

1. Open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. In the navigation pane, choose Your VPCs.
3. Select your VPC, and then select Enable ClassicLink from the Actions list.
4. In the confirmation dialog box, choose Yes, Enable.

Step 4: Create an AMI from Your EC2-Classic Instance

An AMI is a template for launching your instance. You can create your own AMI based on an existing EC2-Classic instance, then use that AMI to launch instances into your VPC.

For more information, see Creating a Custom Windows AMI (p. 64).

(Optional) Store Your Data on Amazon EBS Volumes

You can create an Amazon EBS volume and use it to back up and store the data on your instance—like you would use a physical hard drive. Amazon EBS volumes can be attached and detached from any instance in the same Availability Zone. You can detach a volume from your instance in EC2-Classic, and attach it to a new instance that you launch into your VPC in the same Availability Zone.

For more information about Amazon EBS volumes, see the following topics:

- Amazon EBS Volumes (p. 783)
- Creating an Amazon EBS Volume (p. 800)
- Attaching an Amazon EBS Volume to an Instance (p. 803)
To back up the data on your Amazon EBS volume, you can take periodic snapshots of your volume. If you need to, you can restore an Amazon EBS volume from your snapshot. For more information about Amazon EBS snapshots, see the following topics:

- Amazon EBS Snapshots (p. 818)
- Creating Amazon EBS Snapshots (p. 822)
- Restoring an Amazon EBS Volume from a Snapshot (p. 802)

**Step 5: Launch an Instance Into Your VPC**

The next step in the migration process is to launch instances into your VPC so that you can start transferring functionality to them. You can use the AMIs that you created in the previous step to launch instances into your VPC. The instances will have the same data and configurations as your existing EC2-Classic instances.

**To launch an instance into your VPC using your custom AMI**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. On the dashboard, choose **Launch Instance**.
3. On the **Choose an Amazon Machine Image** page, select the **My AMIs** category, and select the AMI you created.
4. On the **Choose an Instance Type** page, select the type of instance, and choose **Next: Configure Instance Details**.
5. On the **Configure Instance Details** page, select your VPC from the **Network** list. Select the required subnet from the **Subnet** list. Configure any other details you require, then go through the next pages of the wizard until you reach the **Configure Security Group** page.
6. Select **Select an existing group**, and select the security group you created earlier. Choose **Review and Launch**.
7. Review your instance details, then choose **Launch** to specify a key pair and launch your instance.

For more information about the parameters you can configure in each step of the wizard, see Launching an Instance Using the Launch Instance Wizard (p. 334).

After you've launched your instance and it's in the **running** state, you can connect to it and configure it as required.

**Step 6: Link Your EC2-Classic Instances to Your VPC**

After you've configured your instances and made the functionality of your application available in the VPC, you can use ClassicLink to enable private IP communication between your new VPC instances and your EC2-Classic instances.

**To link an instance to a VPC**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Select your EC2-Classic instance, then choose **Actions, ClassicLink, and Link to VPC**.
   
   **Note**
   
   Ensure that your instance is in the **running** state.
4. In the dialog box, select your ClassicLink-enabled VPC (only VPCs that are enabled for ClassicLink are displayed).
5. Select one or more of the VPC security groups to associate with your instance. When you are done, choose **Link to VPC**.
Step 7: Complete the VPC Migration

Depending on the size of your application and the functionality that must be migrated, repeat steps 4 to 6 until you've moved all the components of your application from EC2-Classic into your VPC.

After you've enabled internal communication between the EC2-Classic and VPC instances, you must update your application to point to your migrated service in your VPC, instead of your service in the EC2-Classic platform. The exact steps for this depend on your application’s design. Generally, this includes updating your destination IP addresses to point to the IP addresses of your VPC instances instead of your EC2-Classic instances. You can migrate your Elastic IP addresses that you are currently using in the EC2-Classic platform to a VPC. For more information, see Migrating an Elastic IP Address from EC2-Classic (p. 757).

After you've completed this step and you've tested that the application is functioning from your VPC, you can terminate your EC2-Classic instances, and disable ClassicLink for your VPC. You can also clean up any EC2-Classic resources that you may no longer need to avoid incurring charges for them; for example, you can release Elastic IP addresses, and delete the volumes that were associated with your EC2-Classic instances.
Storage

Amazon EC2 provides you with flexible, cost effective, and easy-to-use data storage options for your instances. Each option has a unique combination of performance and durability. These storage options can be used independently or in combination to suit your requirements.

After reading this section, you should have a good understanding about how you can use the data storage options supported by Amazon EC2 to meet your specific requirements. These storage options include the following:

- Amazon Elastic Block Store (p. 782)
- Amazon EC2 Instance Store (p. 917)
- Amazon Simple Storage Service (Amazon S3) (p. 926)

The following figure shows the relationship between these storage options and your instance.

Amazon EBS

Amazon EBS provides durable, block-level storage volumes that you can attach to a running instance. You can use Amazon EBS as a primary storage device for data that requires frequent and granular updates. For example, Amazon EBS is the recommended storage option when you run a database on an instance.

An EBS volume behaves like a raw, unformatted, external block device that you can attach to a single instance. The volume persists independently from the running life of an instance. After an EBS volume is attached to an instance, you can use it like any other physical hard drive. As illustrated in the previous figure, multiple volumes can be attached to an instance. You can also detach an EBS volume from one instance and attach it to another instance. You can dynamically change the configuration of a volume attached to an instance. EBS volumes can also be created as encrypted volumes using the Amazon EBS encryption feature. For more information, see Amazon EBS Encryption (p. 865).

To keep a backup copy of your data, you can create a snapshot of an EBS volume, which is stored in Amazon S3. You can create an EBS volume from a snapshot, and attach it to another instance. For more information, see Amazon Elastic Block Store (p. 782).

Amazon EC2 Instance Store

Many instances can access storage from disks that are physically attached to the host computer. This disk storage is referred to as instance store. Instance store provides temporary block-level storage for
instances. The data on an instance store volume persists only during the life of the associated instance; if you stop or terminate an instance, any data on instance store volumes is lost. For more information, see Amazon EC2 Instance Store (p. 917).

Amazon S3

Amazon S3 provides access to reliable and inexpensive data storage infrastructure. It is designed to make web-scale computing easier by enabling you to store and retrieve any amount of data, at any time, from within Amazon EC2 or anywhere on the web. For example, you can use Amazon S3 to store backup copies of your data and applications. Amazon EC2 uses Amazon S3 to store EBS snapshots and instance store-backed AMIs. For more information, see Amazon Simple Storage Service (Amazon S3) (p. 926).

Adding Storage

Every time you launch an instance from an AMI, a root storage device is created for that instance. The root storage device contains all the information necessary to boot the instance. You can specify storage volumes in addition to the root device volume when you create an AMI or launch an instance using block device mapping. For more information, see Block Device Mapping (p. 930).

You can also attach EBS volumes to a running instance. For more information, see Attaching an Amazon EBS Volume to an Instance (p. 803).

Amazon Elastic Block Store (Amazon EBS)

Amazon Elastic Block Store (Amazon EBS) provides block level storage volumes for use with EC2 instances. EBS volumes behave like raw, unformatted block devices. You can mount these volumes as devices on your instances. You can mount multiple volumes on the same instance, but each volume can be attached to only one instance at a time. You can create a file system on top of these volumes, or use them in any way you would use a block device (like a hard drive). You can dynamically change the configuration of a volume attached to an instance.

EBS volumes are highly available and reliable storage volumes that can be attached to any running instance that is in the same Availability Zone. EBS volumes that are attached to an EC2 instance are exposed as storage volumes that persist independently from the life of the instance. With Amazon EBS, you pay only for what you use. For more information about Amazon EBS pricing, see the Projecting Costs section of the Amazon Elastic Block Store page.

You can attach multiple volumes to the same instance within the limits specified by your AWS account. Your account has a limit on the number of EBS volumes that you can use, and the total storage available to you. For more information about these limits, and how to request an increase in your limits, see Request to Increase the Amazon EBS Volume Limit.

Amazon EBS is recommended when data must be quickly accessible and requires long-term persistence. EBS volumes are particularly well-suited for use as the primary storage for file systems, databases, or for any applications that require fine granular updates and access to raw, unformatted, block-level storage. Amazon EBS is well suited to both database-style applications that rely on random reads and writes, and to throughput-intensive applications that perform long, continuous reads and writes.

Contents

- Features of Amazon EBS (p. 783)
- Amazon EBS Volumes (p. 783)
- Amazon EBS Snapshots (p. 818)
- Amazon EBS Data Services (p. 855)
- Amazon EBS and NVMe on Windows Instances (p. 875)
- Amazon EBS–Optimized Instances (p. 876)
Features of Amazon EBS

- EBS volumes are created in a specific Availability Zone, and can then be attached to any instances in that same Availability Zone. To make a volume available outside of the Availability Zone, you can create a snapshot and restore that snapshot to a new volume anywhere in that Region. You can copy snapshots to other Regions and then restore them to new volumes there, making it easier to leverage multiple AWS Regions for geographical expansion, data center migration, and disaster recovery.

- Amazon EBS provides the following volume types: General Purpose SSD (gp2), Provisioned IOPS SSD (io1), Throughput Optimized HDD (st1), and Cold HDD (sc1). The following is a summary of performance and use cases for each volume type.

  - General Purpose SSD volumes offer a base performance of 3 IOPS/GB, with the ability to burst to 3,000 IOPS for extended periods of time. These volumes are ideal for a broad range of use cases such as boot volumes, small and medium-size databases, and development and test environments. For more information, see General Purpose SSD (gp2) Volumes (p. 788).

  - Provisioned IOPS SSD volumes support up to 64,000 IOPS and 1,000 MiB/s of throughput. This allows you to predictably scale to tens of thousands of IOPS per EC2 instance. For more information, see Provisioned IOPS SSD (io1) Volumes (p. 791).

  - Throughput Optimized HDD volumes provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. These volumes are ideal for large, sequential workloads such as Amazon EMR, ETL, data warehouses, and log processing. For more information, see Throughput Optimized HDD (st1) Volumes (p. 792).

  - Cold HDD volumes provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. These volumes are ideal for large, sequential, cold-data workloads. If you require infrequent access to your data and are looking to save costs, these volumes provide inexpensive block storage. For more information, see Cold HDD (sc1) Volumes (p. 794).

- You can create your EBS volumes as encrypted volumes, in order to meet a wide range of data-at-rest encryption requirements for regulated/audited data and applications. When you create an encrypted EBS volume and attach it to a supported instance type, data stored at rest on the volume, disk I/O, and snapshots created from the volume are all encrypted. The encryption occurs on the servers that host EC2 instances, providing encryption of data-in-transit from EC2 instances to EBS storage. For more information, see Amazon EBS Encryption (p. 865).

- You can create point-in-time snapshots of EBS volumes, which are persisted to Amazon S3. Snapshots protect data for long-term durability, and they can be used as the starting point for new EBS volumes. The same snapshot can be used to instantiate as many volumes as you wish. These snapshots can be copied across AWS Regions. For more information, see Amazon EBS Snapshots (p. 818).

- Performance metrics, such as bandwidth, throughput, latency, and average queue length, are available through the AWS Management Console. These metrics, provided by Amazon CloudWatch, allow you to monitor the performance of your volumes to make sure that you are providing enough performance for your applications without paying for resources you don’t need. For more information, see Amazon EBS Volume Performance on Windows Instances (p. 887).

Amazon EBS Volumes

An Amazon EBS volume is a durable, block-level storage device that you can attach to a single EC2 instance. You can use EBS volumes as primary storage for data that requires frequent updates, such as the system drive for an instance or storage for a database application. You can also use them for throughput-intensive applications that perform continuous disk scans. EBS volumes persist independently from the running life of an EC2 instance.
After a volume is attached to an instance, you can use it like any other physical hard drive. EBS volumes are flexible. For current-generation volumes attached to current-generation instance types, you can dynamically increase size, modify the provisioned IOPS capacity, and change volume type on live production volumes.

Amazon EBS provides the following volume types: General Purpose SSD (gp2), Provisioned IOPS SSD (io1), Throughput Optimized HDD (st1), Cold HDD (sc1), and Magnetic (standard, a previous-generation type). They differ in performance characteristics and price, allowing you to tailor your storage performance and cost to the needs of your applications. For more information, see Amazon EBS Volume Types (p. 786).

### Benefits of Using EBS Volumes

EBS volumes provide several benefits that are not supported by instance store volumes.

- **Data availability**

  When you create an EBS volume in an Availability Zone, it is automatically replicated within that zone to prevent data loss due to failure of any single hardware component. After you create a volume, you can attach it to any EC2 instance in the same Availability Zone. After you attach a volume, it appears as a native block device similar to a hard drive or other physical device. At that point, the instance can interact with the volume just as it would with a local drive. The instance can format the EBS volume with a file system, such as NTFS, and then install applications.

  An EBS volume can be attached to only one instance at a time, but multiple volumes can be attached to a single instance. If you attach multiple volumes to a device that you have named, you can stripe data across the volumes for increased I/O and throughput performance.

  An EBS volume and the instance to which it attaches must be in the same Availability Zone.

  You can get monitoring data for your EBS volumes, including root device volumes for EBS-backed instances, at no additional charge. For more information about monitoring metrics, see Amazon CloudWatch Metrics for Amazon EBS (p. 903). For information about tracking the status of your volumes, see Amazon CloudWatch Events for Amazon EBS (p. 907).

- **Data persistence**

  An EBS volume is off-instance storage that can persist independently from the life of an instance. You continue to pay for the volume usage as long as the data persists.

  EBS volumes that are attached to a running instance can automatically detach from the instance with their data intact when the instance is terminated if you uncheck the **Delete on Termination**
checkbox when you configure EBS volumes for your instance on the EC2 console. The volume can then be reattached to a new instance, enabling quick recovery. If the checkbox for **Delete on Termination** is checked, the volume(s) will delete upon termination of the EC2 instance. If you are using an EBS-backed instance, you can stop and restart that instance without affecting the data stored in the attached volume. The volume remains attached throughout the stop-start cycle. This enables you to process and store the data on your volume indefinitely, only using the processing and storage resources when required. The data persists on the volume until the volume is deleted explicitly. The physical block storage used by deleted EBS volumes is overwritten with zeroes before it is allocated to another account. If you are dealing with sensitive data, you should consider encrypting your data manually or storing the data on a volume protected by Amazon EBS encryption. For more information, see Amazon EBS Encryption (p. 865).

By default, the root EBS volume that is created and attached to an instance at launch is deleted when that instance is terminated. You can modify this behavior by changing the value of the flag `DeleteOnTermination` to `false` when you launch the instance. This modified value causes the volume to persist even after the instance is terminated, and enables you to attach the volume to another instance.

By default, additional EBS volumes that are created and attached to an instance at launch are not deleted when that instance is terminated. You can modify this behavior by changing the value of the flag `DeleteOnTermination` to `true` when you launch the instance. This modified value causes the volumes to be deleted when the instance is terminated.

- **Data encryption**

For simplified data encryption, you can create encrypted EBS volumes with the Amazon EBS encryption feature. All EBS volume types support encryption. You can use encrypted EBS volumes to meet a wide range of data-at-rest encryption requirements for regulated/audited data and applications. Amazon EBS encryption uses 256-bit Advanced Encryption Standard algorithms (AES-256) and an Amazon-managed key infrastructure. The encryption occurs on the server that hosts the EC2 instance, providing encryption of data-in-transit from the EC2 instance to Amazon EBS storage. For more information, see Amazon EBS Encryption (p. 865).

Amazon EBS encryption uses AWS Key Management Service (AWS KMS) master keys when creating encrypted volumes and any snapshots created from your encrypted volumes. The first time you create an encrypted EBS volume in a region, a default master key is created for you automatically. This key is used for Amazon EBS encryption unless you select a customer master key (CMK) that you created separately using AWS KMS. Creating your own CMK gives you more flexibility, including the ability to create, rotate, disable, define access controls, and audit the encryption keys used to protect your data. For more information, see the AWS Key Management Service Developer Guide.

- **Snapshots**

Amazon EBS provides the ability to create snapshots (backups) of any EBS volume and write a copy of the data in the volume to Amazon S3, where it is stored redundantly in multiple Availability Zones. The volume does not need to be attached to a running instance in order to take a snapshot. As you continue to write data to a volume, you can periodically create a snapshot of the volume to use as a baseline for new volumes. These snapshots can be used to create multiple new EBS volumes or move volumes across Availability Zones. Snapshots of encrypted EBS volumes are automatically encrypted.

When you create a new volume from a snapshot, it's an exact copy of the original volume at the time the snapshot was taken. EBS volumes that are restored from encrypted snapshots are automatically encrypted. By optionally specifying a different Availability Zone, you can use this functionality to create a duplicate volume in that zone. The snapshots can be shared with specific AWS accounts or made public. When you create snapshots, you incur charges in Amazon S3 based on the volume's total size. For a successive snapshot of the volume, you are only charged for any additional data beyond the volume's original size.

Snapshots are incremental backups, meaning that only the blocks on the volume that have changed after your most recent snapshot are saved. If you have a volume with 100 GiB of data, but only 5 GiB
of data have changed since your last snapshot, only the 5 GiB of modified data is written to Amazon S3. Even though snapshots are saved incrementally, the snapshot deletion process is designed so that you need to retain only the most recent snapshot in order to restore the volume.

To help categorize and manage your volumes and snapshots, you can tag them with metadata of your choice. For more information, see Tagging Your Amazon EC2 Resources (p. 964).

- Flexibility

EBS volumes support live configuration changes while in production. You can modify volume type, volume size, and IOPS capacity without service interruptions.

**Amazon EBS Volume Types**

Amazon EBS provides the following volume types, which differ in performance characteristics and price, so that you can tailor your storage performance and cost to the needs of your applications. The volumes types fall into two categories:

- SSD-backed volumes optimized for transactional workloads involving frequent read/write operations with small I/O size, where the dominant performance attribute is IOPS.
- HDD-backed volumes optimized for large streaming workloads where throughput (measured in MiB/s) is a better performance measure than IOPS.

There are several factors that can affect the performance of EBS volumes, such as instance configuration, I/O characteristics, and workload demand. For more information about getting the most out of your EBS volumes, see Amazon EBS Volume Performance on Windows Instances (p. 887).

For more information about pricing, see Amazon EBS Pricing.

**Volume Characteristics**

The following table describes the use cases and performance characteristics for each volume type. The default volume type is General Purpose SSD (gp2).

<table>
<thead>
<tr>
<th>Solid-State Drives (SSD)</th>
<th>Hard Disk Drives (HDD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Volume Type</strong></td>
<td><strong>Provisioned IOPS SSD (io1)</strong></td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>General purpose SSD volume that balances price and performance for a wide variety of workloads</td>
</tr>
</tbody>
</table>
| **Use Cases**            | • Recommended for most workloads  
                          • System boot volumes  
                          • Virtual desktops | • Critical business applications that require sustained IOPS performance, or more than 16,000 IOPS or 250 MiB/s of throughput per volume | • Streaming workloads requiring consistent, fast throughput at a low price  
                          • Big data | • Throughput-oriented storage for large volumes of data that is infrequently accessed |
Solid-State Drives (SSD) | Hard Disk Drives (HDD)
--- | ---
• Low-latency interactive apps | • Large database workloads, such as:
• Development and test environments | • MongoDB
| Cassandra
| Microsoft SQL Server
| MySQL
| PostgreSQL
| Oracle
| Data warehouses
| Log processing
| Cannot be a boot volume
| Scenarios where the lowest storage cost is important
| Cannot be a boot volume

<table>
<thead>
<tr>
<th>API Name</th>
<th>Volume Size</th>
<th>Max IOPS per Volume</th>
<th>Max Throughput per Volume</th>
<th>Max IOPS per Instance ††</th>
<th>Max Throughput per Instance ††</th>
<th>Dominant Performance Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp2</td>
<td>1 GiB - 16 TiB</td>
<td>16,000 (16 KiB I/O) *</td>
<td>250 MiB/s *</td>
<td>80,000</td>
<td>1,750 MiB/s</td>
<td>IOPS</td>
</tr>
<tr>
<td>io1</td>
<td>4 GiB - 16 TiB</td>
<td>64,000 (16 KiB I/O) †</td>
<td>1,000 MiB/s †</td>
<td>80,000</td>
<td>1,750 MiB/s</td>
<td>IOPS</td>
</tr>
<tr>
<td>st1</td>
<td>500 GiB - 16 TiB</td>
<td>500 (1 MiB I/O)</td>
<td>500 MiB/s</td>
<td>80,000</td>
<td>1,750 MiB/s</td>
<td>MiB/s</td>
</tr>
<tr>
<td>sc1</td>
<td>500 GiB - 16 TiB</td>
<td>250 (1 MiB I/O)</td>
<td>250 MiB/s</td>
<td>80,000</td>
<td>1,750 MiB/s</td>
<td>MiB/s</td>
</tr>
</tbody>
</table>

* The throughput limit is between 128 MiB/s and 250 MiB/s, depending on the volume size. Volumes smaller than 170 GiB deliver a maximum throughput of 128 MiB/s. Volumes larger than 170 GiB but smaller than 334 GiB deliver a maximum throughput of 250 MiB/s if burst credits are available. Volumes larger than or equal to 334 GiB deliver 250 MiB/s regardless of burst credits. Older gp2 volumes might not reach full performance unless you modify the volume. For more information, see Amazon EBS Elastic Volumes (p. 855).

† Maximum IOPS and throughput are guaranteed only on Nitro-based Instances (p. 122). Other instances guarantee up to 32,000 IOPS and 500 MiB/s. Older io1 volumes might not reach full performance unless you modify the volume. For more information, see Amazon EBS Elastic Volumes (p. 855).

†† To achieve this throughput, you must have an instance that supports EBS optimization (p. 876).

**Previous Generation Volume Types**

The following table describes previous-generation EBS volume types. If you need higher performance or performance consistency than previous-generation volumes can provide, we recommend that you consider using General Purpose SSD (gp2) or other current volume types. For more information, see Previous Generation Volumes.
Hard Disk Drives (HDD)

<table>
<thead>
<tr>
<th>Use Cases</th>
<th>Workloads where data is infrequently accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>API Name</td>
<td>standard</td>
</tr>
<tr>
<td>Volume Size</td>
<td>1 GiB-1 TiB</td>
</tr>
<tr>
<td>Max IOPS per Volume</td>
<td>40–200</td>
</tr>
<tr>
<td>Max Throughput per Volume</td>
<td>40–90 MiB/s</td>
</tr>
<tr>
<td>Max IOPS per Instance</td>
<td>80,000</td>
</tr>
<tr>
<td>Max Throughput per Instance</td>
<td>1,750 MiB/s</td>
</tr>
<tr>
<td>Dominant Performance Attribute</td>
<td>IOPS</td>
</tr>
</tbody>
</table>

General Purpose SSD (gp2) Volumes

General Purpose SSD (gp2) volumes offer cost-effective storage that is ideal for a broad range of workloads. These volumes deliver single-digit millisecond latencies and the ability to burst to 3,000 IOPS for extended periods of time. Between a minimum of 100 IOPS (at 33.33 GiB and below) and a maximum of 16,000 IOPS (at 5,334 GiB and above), baseline performance scales linearly at 3 IOPS per GiB of volume size. AWS designs gp2 volumes to deliver 90% of the provisioned performance 99% of the time. A gp2 volume can range in size from 1 GiB to 16 TiB.

I/O Credits and Burst Performance

The performance of gp2 volumes is tied to volume size, which determines the baseline performance level of the volume and how quickly it accumulates I/O credits; larger volumes have higher baseline performance levels and accumulate I/O credits faster. I/O credits represent the available bandwidth that your gp2 volume can use to burst large amounts of I/O when more than the baseline performance is needed. The more credits your volume has for I/O, the more time it can burst beyond its baseline performance level and the better it performs when more performance is needed. The following diagram shows the burst-bucket behavior for gp2.

```
GP2 burst bucket
```

Each volume receives an initial I/O credit balance of 5.4 million I/O credits, which is enough to sustain the maximum burst performance of 3,000 IOPS for 30 minutes. This initial credit balance is designed to provide a fast initial boot cycle for boot volumes and to provide a good bootstrapping experience.
Volumes earn I/O credits at the baseline performance rate of 3 IOPS per GiB of volume size. For example, a 100 GiB gp2 volume has a baseline performance of 300 IOPS.

When your volume requires more than the baseline performance I/O level, it draws on I/O credits in the credit balance to burst to the required performance level, up to a maximum of 3,000 IOPS. Volumes larger than 1,000 GiB have a baseline performance that is equal or greater than the maximum burst performance, and their I/O credit balance never depletes. When your volume uses fewer I/O credits than it earns in a second, unused I/O credits are added to the I/O credit balance. The maximum I/O credit balance for a volume is equal to the initial credit balance (5.4 million I/O credits).

**Note**
For a volume 1 TiB or larger, baseline performance is higher than maximum burst performance, so I/O credits are never spent. If the volume is attached to a Nitro-based instance, the reported burst balance is 0%. For a non-Nitro-based instance, the reported burst balance is 100%.

The following table lists several volume sizes and the associated baseline performance of the volume (which is also the rate at which it accumulates I/O credits), the burst duration at the 3,000 IOPS maximum (when starting with a full credit balance), and the time in seconds that the volume would take to refill an empty credit balance.

<table>
<thead>
<tr>
<th>Volume size (GiB)</th>
<th>Baseline performance (IOPS)</th>
<th>Burst duration when driving sustained 3,000 IOPS (second)</th>
<th>Seconds to fill empty credit balance when driving no IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>1802</td>
<td>54,000</td>
</tr>
<tr>
<td>100</td>
<td>300</td>
<td>2,000</td>
<td>18,000</td>
</tr>
<tr>
<td>250</td>
<td>750</td>
<td>2,400</td>
<td>7,200</td>
</tr>
<tr>
<td>334 (Min. size for max throughput)</td>
<td>1002</td>
<td>2703</td>
<td>5389</td>
</tr>
<tr>
<td>500</td>
<td>1,500</td>
<td>3,600</td>
<td>3,600</td>
</tr>
<tr>
<td>750</td>
<td>2,250</td>
<td>7,200</td>
<td>2,400</td>
</tr>
<tr>
<td>1,000</td>
<td>3,000</td>
<td>N/A*</td>
<td>N/A*</td>
</tr>
<tr>
<td>5,334 (Min. size for max IOPS)</td>
<td>16,000</td>
<td>N/A*</td>
<td>N/A*</td>
</tr>
</tbody>
</table>
**Volume size (GiB)** | **Baseline performance (IOPS)** | **Burst duration when driving sustained 3,000 IOPS (second)** | **Seconds to fill empty credit balance when driving no IO**
---|---|---|---
16,384 (16 TiB, max volume size) | 16,000 | N/A* | N/A*

* Bursting and I/O credits are only relevant to volumes under 1,000 GiB, where burst performance exceeds baseline performance.

The burst duration of a volume is dependent on the size of the volume, the burst IOPS required, and the credit balance when the burst begins. This is shown in the following equation:

\[
\text{Burst duration} = \frac{\text{Credit balance}}{(\text{Burst IOPS}) - 3(\text{Volume size in GiB})}
\]

**What happens if I empty my I/O credit balance?**

If your gp2 volume uses all of its I/O credit balance, the maximum IOPS performance of the volume remains at the baseline IOPS performance level (the rate at which your volume earns credits) and the volume's maximum throughput is reduced to the baseline IOPS multiplied by the maximum I/O size. Throughput can never exceed 250 MiB/s. When I/O demand drops below the baseline level and unused credits are added to the I/O credit balance, the maximum IOPS performance of the volume again exceeds the baseline. For example, a 100 GiB gp2 volume with an empty credit balance has a baseline performance of 300 IOPS and a throughput limit of 75 MiB/s (300 I/O operations per second * 256 KiB per I/O operation = 75 MiB/s). The larger a volume is, the greater the baseline performance is and the faster it replenishes the credit balance. For more information about how IOPS are measured, see I/O Characteristics and Monitoring (p. 889).

If you notice that your volume performance is frequently limited to the baseline level (due to an empty I/O credit balance), you should consider using a larger gp2 volume (with a higher baseline performance level) or switching to an io1 volume for workloads that require sustained IOPS performance greater than 16,000 IOPS.

For information about using CloudWatch metrics and alarms to monitor your burst bucket balance, see Monitoring the Burst Bucket Balance for gp2, st1, and sc1 Volumes (p. 798).

**Throughput Performance**

Throughput for a gp2 volume can be calculated using the following formula, up to the throughput limit of 250 MiB/s:

\[
\text{Throughput in MiB/s} = ((\text{Volume size in GiB}) \times (\text{IOPS per GiB}) \times (\text{I/O size in KiB}))
\]

Assuming \(V = \text{volume size}, I = \text{I/O size}, R = \text{I/O rate}, \) and \(T = \text{throughput}, \) this can be simplified to:

\[
T = VIR
\]

The smallest volume size that achieves the maximum throughput is given by:

\[
V = \frac{T}{IR}
\]
Provisioned IOPS SSD (io1) Volumes

Provisioned IOPS SSD (io1) volumes are designed to meet the needs of I/O-intensive workloads, particularly database workloads, that are sensitive to storage performance and consistency. Unlike gp2, which uses a bucket and credit model to calculate performance, an io1 volume allows you to specify a consistent IOPS rate when you create the volume, and Amazon EBS delivers the provisioned IOPS performance 99.9 percent of the time.

An io1 volume can range in size from 4 GiB to 16 TiB. You can provision from 100 IOPS up to 64,000 IOPS per volume on Nitro-based Instances (p. 122) instances and up to 32,000 on other instances. The maximum ratio of provisioned IOPS to requested volume size (in GiB) is 50:1. For example, a 100 GiB volume can be provisioned with up to 5,000 IOPS. On a supported instance type, any volume 1,280 GiB in size or greater allows provisioning up to the 64,000 IOPS maximum (50 x 1,280 GiB = 64,000).

An io1 volume provisioned with up to 32,000 IOPS supports a maximum I/O size of 256 KiB and yields as much as 500 MiB/s of throughput. With the I/O size at the maximum, peak throughput is reached at 2,000 IOPS. A volume provisioned with more than 32,000 IOPS (up to the cap of 64,000 IOPS) supports a maximum I/O size of 16 KiB and yields as much as 1,000 MiB/s of throughput. The following graph illustrates these performance characteristics:

Your per-I/O latency experience depends on the IOPS provisioned and your workload pattern. For the best per-I/O latency experience, we recommend that you provision an IOPS-to-GiB ratio greater than 2:1. For example, a 2,000 IOPS volume should be smaller than 1,000 GiB.

Note
Some AWS accounts created before 2012 might have access to Availability Zones in us-west-1 or ap-northeast-1 that do not support Provisioned IOPS SSD (io1) volumes. If you are unable
to create an io1 volume (or launch an instance with an io1 volume in its block device mapping) in one of these Regions, try a different Availability Zone in the Region. You can verify that an Availability Zone supports io1 volumes by creating a 4 GiB io1 volume in that zone.

**Throughput Optimized HDD (st1) Volumes**

Throughput Optimized HDD (st1) volumes provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. This volume type is a good fit for large, sequential workloads such as Amazon EMR, ETL, data warehouses, and log processing. Bootable st1 volumes are not supported.

Throughput Optimized HDD (st1) volumes, though similar to Cold HDD (sc1) volumes, are designed to support frequently accessed data.

This volume type is optimized for workloads involving large, sequential I/O, and we recommend that customers with workloads performing small, random I/O use gp2. For more information, see Inefficiency of Small Read/Writes on HDD (p. 797).

**Throughput Credits and Burst Performance**

Like gp2, st1 uses a burst-bucket model for performance. Volume size determines the baseline throughput of your volume, which is the rate at which the volume accumulates throughput credits. Volume size also determines the burst throughput of your volume, which is the rate at which you can spend credits when they are available. Larger volumes have higher baseline and burst throughput. The more credits your volume has, the longer it can drive I/O at the burst level.

The following diagram shows the burst-bucket behavior for st1.

**ST1 burst bucket**

Subject to throughput and throughput-credit caps, the available throughput of an st1 volume is expressed by the following formula:

\[(\text{Volume size}) \times (\text{Credit accumulation rate per TiB}) = \text{Throughput}\]

For a 1-TiB st1 volume, burst throughput is limited to 250 MiB/s, the bucket fills with credits at 40 MiB/s, and it can hold up to 1 TiB-worth of credits.

Larger volumes scale these limits linearly, with throughput capped at a maximum of 500 MiB/s. After the bucket is depleted, throughput is limited to the baseline rate of 40 MiB/s per TiB.

On volume sizes ranging from 0.5 to 16 TiB, baseline throughput varies from 20 to a cap of 500 MiB/s, which is reached at 12.5 TiB as follows:

\[40 \text{ MiB/s} \times 12.5 \text{ TiB} = 500 \text{ MiB/s}\]
**1 TiB**

Burst throughput varies from 125 MiB/s to a cap of 500 MiB/s, which is reached at 2 TiB as follows:

\[
\text{250 MiB/s} \\
\frac{2 \text{ TiB} \times \frac{250 \text{ MiB/s}}{1 \text{ TiB}}}{1 \text{ TiB}} = 500 \text{ MiB/s}
\]

The following table states the full range of base and burst throughput values for st1:

<table>
<thead>
<tr>
<th>Volume Size (TiB)</th>
<th>ST1 Base Throughput (MiB/s)</th>
<th>ST1 Burst Throughput (MiB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>20</td>
<td>125</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>250</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>500</td>
</tr>
<tr>
<td>3</td>
<td>120</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>160</td>
<td>500</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
<td>500</td>
</tr>
<tr>
<td>6</td>
<td>240</td>
<td>500</td>
</tr>
<tr>
<td>7</td>
<td>280</td>
<td>500</td>
</tr>
<tr>
<td>8</td>
<td>320</td>
<td>500</td>
</tr>
<tr>
<td>9</td>
<td>360</td>
<td>500</td>
</tr>
<tr>
<td>10</td>
<td>400</td>
<td>500</td>
</tr>
<tr>
<td>11</td>
<td>440</td>
<td>500</td>
</tr>
<tr>
<td>12</td>
<td>480</td>
<td>500</td>
</tr>
<tr>
<td>12.5</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>13</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>14</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>15</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>16</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

The following diagram plots the table values:
Note
When you create a snapshot of a Throughput Optimized HDD (st1) volume, performance may drop as far as the volume's baseline value while the snapshot is in progress.

For information about using CloudWatch metrics and alarms to monitor your burst bucket balance, see Monitoring the Burst Bucket Balance for gp2, st1, and sc1 Volumes (p. 798).

Cold HDD (sc1) Volumes

Cold HDD (sc1) volumes provide low-cost magnetic storage that defines performance in terms of throughput rather than IOPS. With a lower throughput limit than st1, sc1 is a good fit for large, sequential cold-data workloads. If you require infrequent access to your data and are looking to save costs, sc1 provides inexpensive block storage. Bootable sc1 volumes are not supported.

Cold HDD (sc1) volumes, though similar to Throughput Optimized HDD (st1) volumes, are designed to support infrequently accessed data.

Note
This volume type is optimized for workloads involving large, sequential I/O, and we recommend that customers with workloads performing small, random I/O use gp2. For more information, see Inefficiency of Small Read/Writes on HDD (p. 797).

Throughput Credits and Burst Performance

Like gp2, sc1 uses a burst-bucket model for performance. Volume size determines the baseline throughput of your volume, which is the rate at which the volume accumulates throughput credits. Volume size also determines the burst throughput of your volume, which is the rate at which you can spend credits when they are available. Larger volumes have higher baseline and burst throughput. The more credits your volume has, the longer it can drive I/O at the burst level.

SC1 burst bucket

Subject to throughput and throughput-credit caps, the available throughput of an sc1 volume is expressed by the following formula:

\[(\text{Volume size}) \times (\text{Credit accumulation rate per TiB}) = \text{Throughput}\]

For a 1-TiB sc1 volume, burst throughput is limited to 80 MiB/s, the bucket fills with credits at 12 MiB/s, and it can hold up to 1 TiB-worth of credits.

Larger volumes scale these limits linearly, with throughput capped at a maximum of 250 MiB/s. After the bucket is depleted, throughput is limited to the baseline rate of 12 MiB/s per TiB.

On volume sizes ranging from 0.5 to 16 TiB, baseline throughput varies from 6 MiB/s to a maximum of 192 MiB/s, which is reached at 16 TiB as follows:

\[
\begin{align*}
12 \text{ MiB/s} \\
16 \text{ TiB} \times 12 \text{ MiB/s} &= 192 \text{ MiB/s}
\end{align*}
\]
Burst throughput varies from 40 MiB/s to a cap of 250 MiB/s, which is reached at 3.125 TiB as follows:

\[
\frac{80 \text{ MiB/s}}{3.125 \text{ TiB}} = \frac{-}{1 \text{ TiB}} = 250 \text{ MiB/s}
\]

The following table states the full range of base and burst throughput values for sc1:

<table>
<thead>
<tr>
<th>Volume Size (TiB)</th>
<th>SC1 Base Throughput (MiB/s)</th>
<th>SC1 Burst Throughput (MiB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>6</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>160</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>240</td>
</tr>
<tr>
<td>3.125</td>
<td>37.5</td>
<td>250</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>250</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>250</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>250</td>
</tr>
<tr>
<td>7</td>
<td>84</td>
<td>250</td>
</tr>
<tr>
<td>8</td>
<td>96</td>
<td>250</td>
</tr>
<tr>
<td>9</td>
<td>108</td>
<td>250</td>
</tr>
<tr>
<td>10</td>
<td>120</td>
<td>250</td>
</tr>
<tr>
<td>11</td>
<td>132</td>
<td>250</td>
</tr>
<tr>
<td>12</td>
<td>144</td>
<td>250</td>
</tr>
<tr>
<td>13</td>
<td>156</td>
<td>250</td>
</tr>
<tr>
<td>14</td>
<td>168</td>
<td>250</td>
</tr>
<tr>
<td>15</td>
<td>180</td>
<td>250</td>
</tr>
<tr>
<td>16</td>
<td>192</td>
<td>250</td>
</tr>
</tbody>
</table>

The following diagram plots the table values:
**Note**
When you create a snapshot of a Cold HDD (sc1) volume, performance may drop as far as the volume's baseline value while the snapshot is in progress.

For information about using CloudWatch metrics and alarms to monitor your burst bucket balance, see Monitoring the Burst Bucket Balance for gp2, st1, and sc1 Volumes (p. 798).

**Magnetic (standard)**

Magnetic volumes are backed by magnetic drives and are suited for workloads where data is accessed infrequently, and scenarios where low-cost storage for small volume sizes is important. These volumes deliver approximately 100 IOPS on average, with burst capability of up to hundreds of IOPS, and they can range in size from 1 GiB to 1 TiB.

**Note**
Magnetic is a Previous Generation Volume. For new applications, we recommend using one of the newer volume types. For more information, see Previous Generation Volumes.

For information about using CloudWatch metrics and alarms to monitor your burst bucket balance, see Monitoring the Burst Bucket Balance for gp2, st1, and sc1 Volumes (p. 798).

**Performance Considerations When Using HDD Volumes**

For optimal throughput results using HDD volumes, plan your workloads with the following considerations in mind.

**Throughput Optimized HDD vs. Cold HDD**

The st1 and sc1 bucket sizes vary according to volume size, and a full bucket contains enough tokens for a full volume scan. However, larger st1 and sc1 volumes take longer for the volume scan to complete due to per-instance and per-volume throughput limits. Volumes attached to smaller instances are limited to the per-instance throughput rather than the st1 or sc1 throughput limits.

Both st1 and sc1 are designed for performance consistency of 90% of burst throughput 99% of the time. Non-compliant periods are approximately uniformly distributed, targeting 99% of expected total throughput each hour.

The following table shows ideal scan times for volumes of various size, assuming full buckets and sufficient instance throughput.

In general, scan times are expressed by this formula:

\[
\frac{\text{Volume size}}{\text{Throughput}} = \text{Scan time}
\]

For example, taking the performance consistency guarantees and other optimizations into account, an st1 customer with a 5-TiB volume can expect to complete a full volume scan in 2.91 to 3.27 hours.

<table>
<thead>
<tr>
<th>Volume size</th>
<th>Throughput</th>
<th>Scan time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 TiB</td>
<td>5 TiB</td>
<td>10,486 s</td>
</tr>
<tr>
<td>0.00047684 TiB/s</td>
<td>0.00047684 TiB/s</td>
<td>2.91 hours</td>
</tr>
</tbody>
</table>

2.91 hours + \(0.90)(0.99)\ = 3.27 hours (minimum expected)
Similarly, an sc1 customer with a 5-TiB volume can expect to complete a full volume scan in 5.83 to 6.54 hours.

\[
\frac{5 \text{ TiB}}{0.000238418 \text{ TiB/s}} = 20972 \text{ s} = 5.83 \text{ hours (optimal)}
\]

\[
\frac{5.83 \text{ hours}}{(0.90)(0.99)} = 6.54 \text{ hours (minimum expected)}
\]

<table>
<thead>
<tr>
<th>Volume Size (TiB)</th>
<th>ST1 Scan Time with Burst (Hours)*</th>
<th>SC1 Scan Time with Burst (Hours)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.17</td>
<td>3.64</td>
</tr>
<tr>
<td>2</td>
<td>1.17</td>
<td>3.64</td>
</tr>
<tr>
<td>3</td>
<td>1.75</td>
<td>3.64</td>
</tr>
<tr>
<td>4</td>
<td>2.33</td>
<td>4.66</td>
</tr>
<tr>
<td>5</td>
<td>2.91</td>
<td>5.83</td>
</tr>
<tr>
<td>6</td>
<td>3.50</td>
<td>6.99</td>
</tr>
<tr>
<td>7</td>
<td>4.08</td>
<td>8.16</td>
</tr>
<tr>
<td>8</td>
<td>4.66</td>
<td>9.32</td>
</tr>
<tr>
<td>9</td>
<td>5.24</td>
<td>10.49</td>
</tr>
<tr>
<td>10</td>
<td>5.83</td>
<td>11.65</td>
</tr>
<tr>
<td>11</td>
<td>6.41</td>
<td>12.82</td>
</tr>
<tr>
<td>12</td>
<td>6.99</td>
<td>13.98</td>
</tr>
<tr>
<td>13</td>
<td>7.57</td>
<td>15.15</td>
</tr>
<tr>
<td>14</td>
<td>8.16</td>
<td>16.31</td>
</tr>
<tr>
<td>15</td>
<td>8.74</td>
<td>17.48</td>
</tr>
<tr>
<td>16</td>
<td>9.32</td>
<td>18.64</td>
</tr>
</tbody>
</table>

* These scan times assume an average queue depth (rounded to the nearest whole number) of four or more when performing 1 MiB of sequential I/O.

Therefore if you have a throughput-oriented workload that needs to complete scans quickly (up to 500 MiB/s), or requires several full volume scans a day, use st1. If you are optimizing for cost, your data is relatively infrequently accessed, and you don't need more than 250 MiB/s of scanning performance, then use sc1.

**Inefficiency of Small Read/Writes on HDD**

The performance model for st1 and sc1 volumes is optimized for sequential I/Os, favoring high-throughput workloads, offering acceptable performance on workloads with mixed IOPS and throughput, and discouraging workloads with small, random I/O.
For example, an I/O request of 1 MiB or less counts as a 1 MiB I/O credit. However, if the I/Os are sequential, they are merged into 1 MiB I/O blocks and count only as a 1 MiB I/O credit.

**Limitations on per-Instance Throughput**

Throughput for st1 and sc1 volumes is always determined by the smaller of the following:

- Throughput limits of the volume
- Throughput limits of the instance

As for all Amazon EBS volumes, we recommend that you select an appropriate EBS-optimized EC2 instance in order to avoid network bottlenecks. For more information, see Amazon EBS–Optimized Instances (p. 876).

**Monitoring the Burst Bucket Balance for gp2, st1, and sc1 Volumes**

You can monitor the burst-bucket level for gp2, st1, and sc1 volumes using the EBS BurstBalance metric available in Amazon CloudWatch. This metric shows the percentage of I/O credits (for gp2) or throughput credits (for st1 and sc1) remaining in the burst bucket. For more information about the BurstBalance metric and other metrics related to I/O, see I/O Characteristics and Monitoring (p. 889). CloudWatch also allows you to set an alarm that notifies you when the BurstBalance value falls to a certain level. For more information, see Creating Amazon CloudWatch Alarms.

**Constraints on the Size and Configuration of an EBS Volume**

The size of an Amazon EBS volume is constrained by the physics and arithmetic of block data storage, as well as by the implementation decisions of operating system (OS) and file system designers. AWS imposes additional limits on volume size to safeguard the reliability of its services.

The following sections describe the most important factors that limit the usable size of an EBS volume and offer recommendations for configuring your EBS volumes.

**Contents**

- Storage Capacity (p. 798)
- Service Limitations (p. 799)
- Partitioning Schemes (p. 799)
- Data Block Sizes (p. 800)

**Storage Capacity**

The following table summarizes the theoretical and implemented storage capacities for the most commonly used file systems on Amazon EBS, assuming a 4,096 byte block size.

<table>
<thead>
<tr>
<th>Partitioning Scheme</th>
<th>Max addressable blocks</th>
<th>Theoretical max size (blocks × block size)</th>
<th>Ext4 implemented max size*</th>
<th>XFS implemented max size**</th>
<th>NTFS implemented max size</th>
<th>Max supported by EBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBR</td>
<td>$2^{32}$</td>
<td>2 TiB</td>
<td>2 TiB</td>
<td>2 TiB</td>
<td>2 TiB</td>
<td>2 TiB</td>
</tr>
<tr>
<td>GPT</td>
<td>$2^{64}$</td>
<td>$8 , \text{ZiB} = 8 \times 1024^3 , \text{TiB}$</td>
<td>1 EiB $= 1024^4 , \text{TiB}$</td>
<td>500 TiB</td>
<td>256 TiB</td>
<td>16 TiB</td>
</tr>
</tbody>
</table>

(extended on RHEL7)
Amazon Elastic Compute Cloud
User Guide for Windows Instances
EBS Volumes

<table>
<thead>
<tr>
<th>Partitioning Scheme</th>
<th>Max addressable blocks</th>
<th>Theoretical max size (blocks × block size)</th>
<th>Ext4 implemented max size*</th>
<th>XFS implemented max size**</th>
<th>NTFS implemented max size</th>
<th>Max supported by EBS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(50 TiB certified on RHEL7)</td>
</tr>
</tbody>
</table>

** https://access.redhat.com/solutions/1532

Service Limitations

Amazon EBS abstracts the massively distributed storage of a data center into virtual hard disk drives. To an operating system installed on an EC2 instance, an attached EBS volume appears to be a physical hard disk drive containing 512-byte disk sectors. The OS manages the allocation of data blocks (or clusters) onto those virtual sectors through its storage management utilities. The allocation is in conformity with a volume partitioning scheme, such as master boot record (MBR) or GUID partition table (GPT), and within the capabilities of the installed file system (ext4, NTFS, and so on).

EBS is not aware of the data contained in its virtual disk sectors; it only ensures the integrity of the sectors. This means that AWS actions and OS actions are independent of each other. When you are selecting a volume size, be aware of the capabilities and limits of both, as in the following cases:

- EBS currently supports a maximum volume size of 16 TiB. This means that you can create an EBS volume as large as 16 TiB, but whether the OS recognizes all of that capacity depends on its own design characteristics and on how the volume is partitioned.
- Amazon EC2 requires Windows boot volumes to use MBR partitioning. As discussed in Partitioning Schemes (p. 799), this means that boot volumes cannot be larger than 2 TiB. Windows data volumes are not subject to this limitation and can use GPT partitioning. If a Windows boot volume that is 2 TiB or larger is converted to use a dynamic MBR partition table, you will see an error for the volume in Disk Manager.
- Windows non-boot volumes that are 2 TiB (2048 GiB) or larger must use a GPT partition table to access the entire volume. If an EBS volume over 2 TiB in size is attached to a Windows instance at launch, it is automatically formatted with a GPT partition table. If you attach an EBS volume over 2 TiB in size to a Windows instance after launch, you must initialize it with a GPT table manually. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804).

Partitioning Schemes

Among other impacts, the partitioning scheme determines how many logical data blocks can be uniquely addressed in a single volume. For more information, see Data Block Sizes (p. 800). The common partitioning schemes in use are master boot record (MBR) and GUID partition table (GPT). The important differences between these schemes can be summarized as follows.

MBR

MBR uses a 32-bit data structure to store block addresses. This means that each data block is mapped with one of $2^{32}$ possible integers. The maximum addressable size of a volume is given by:

\[(2^{32} - 1) \times \text{Block size} = \text{Number of addressable blocks}\]

The block size for MBR volumes is conventionally limited to 512 bytes. Therefore:
### Amazon Elastic Compute Cloud

**User Guide for Windows Instances**

**EBS Volumes**

\[(2^{32} - 1) \times 512 \text{ bytes} = 2 \text{ TiB} - 512 \text{ bytes}\]

Engineering workarounds to increase this 2-TiB limit for MBR volumes have not met with widespread industry adoption. Consequently, Linux and Windows never detect an MBR volume as being larger than 2 TiB even if AWS shows its size to be larger.

**GPT**

GPT uses a 64-bit data structure to store block addresses. This means that each data block is mapped with one of \(2^{64}\) possible integers. The maximum addressable size of a volume is given by:

\[(2^{64} - 1) \times \text{Block size} = \text{Number of addressable blocks}\]

The block size for GPT volumes is commonly 4,096 bytes. Therefore:

\[(2^{64} - 1) \times 4,096 \text{ bytes} = 8 \text{ ZiB} - 4,096 \text{ bytes} = 8 \text{ billion TiB} - 4,096 \text{ bytes}\]

Real-world computer systems don't support anything close to this theoretical maximum. Implemented file-system size is currently limited to 50 TiB for ext4 and 256 TiB for NTFS—both of which exceed the 16-TiB limit imposed by AWS.

**Data Block Sizes**

Data storage on a modern hard drive is managed through *logical block addressing*, an abstraction layer that allows the operating system to read and write data in logical blocks without knowing much about the underlying hardware. The OS relies on the storage device to map the blocks to its physical sectors. EBS advertises 512-byte sectors to the operating system, which reads and writes data to disk using data blocks that are a multiple of the sector size.

The industry default size for logical data blocks is currently 4,096 bytes (4 KiB). Because certain workloads benefit from a smaller or larger block size, file systems support non-default block sizes that can be specified during formatting. Scenarios in which non-default block sizes should be used are outside the scope of this topic, but the choice of block size has consequences for the storage capacity of the volume. The following table shows storage capacity as a function of block size:

<table>
<thead>
<tr>
<th>Block size</th>
<th>Max volume size</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 KiB (default)</td>
<td>16 TiB</td>
</tr>
<tr>
<td>8 KiB</td>
<td>32 TiB</td>
</tr>
<tr>
<td>16 KiB</td>
<td>64 TiB</td>
</tr>
<tr>
<td>32 KiB</td>
<td>128 TiB</td>
</tr>
<tr>
<td>64 KiB (maximum)</td>
<td>256 TiB</td>
</tr>
</tbody>
</table>

The EBS-imposed limit on volume size (16 TiB) is currently equal to the maximum size enabled by 4-KiB data blocks.

**Creating an Amazon EBS Volume**

You can create an Amazon EBS volume that you can then attach to any EC2 instance within the same Availability Zone. You can choose to create an encrypted EBS volume, but encrypted volumes can only be attached to supported instance types. For more information, see Supported Instance Types (p. 866).
If you are creating a volume for a high-performance storage scenario, you should make sure to use a Provisioned IOPS SSD (io1) volume and attach it to an instance with enough bandwidth to support your application, such as an EBS-optimized instance or an instance with 10-Gigabit network connectivity. The same advice holds for Throughput Optimized HDD (st1) and Cold HDD (sc1) volumes. For more information, see Amazon EC2 Instance Configuration (p. 888).

New EBS volumes receive their maximum performance the moment that they are available and do not require initialization (formerly known as pre-warming). However, storage blocks on volumes that were restored from snapshots must be initialized (pulled down from Amazon S3 and written to the volume) before you can access the block. This preliminary action takes time and can cause a significant increase in the latency of an I/O operation the first time each block is accessed. For most applications, amortizing this cost over the lifetime of the volume is acceptable. Performance is restored after the data is accessed once. For more information, see Initializing Amazon EBS Volumes (p. 891).

Methods of Creating a Volume

• You can create an EBS volume and attach it to a running instance. For more information, see the procedure below.
• You can create and attach EBS volumes when you launch instances by specifying a block device mapping. For more information, see Launching an Instance Using the Launch Instance Wizard (p. 334) and Block Device Mapping (p. 930).
• You can restore volumes from previously created snapshots. For more information, see Restoring an Amazon EBS Volume from a Snapshot (p. 802).

To create a new (empty) EBS volume using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region in which you would like to create your volume. This choice is important because some Amazon EC2 resources can be shared between Regions, while others can't. For more information, see Resource Locations (p. 955).
3. In the navigation pane, choose ELASTIC BLOCK STORE, Volumes.
4. Choose Create Volume.
5. For Volume Type, choose a volume type. For more information, see Amazon EBS Volume Types (p. 786).
6. For Size (GiB), type the size of the volume. For more information, see Constraints on the Size and Configuration of an EBS Volume (p. 798).
7. With a Provisioned IOPS SSD volume, for IOPS, type the maximum number of input/output operations per second (IOPS) that the volume should support.
8. For Availability Zone, choose the Availability Zone in which to create the volume. EBS volumes can only be attached to EC2 instances within the same Availability Zone.
9. (Optional) If the instance type supports EBS encryption and you want to encrypt the volume, select Encrypt this volume and choose a CMK. If encryption by default is enabled in this Region, EBS encryption is enabled and the default CMK for EBS encryption is chosen. You can choose a different CMK from Master Key or paste the full ARN of any key that you can access. For more information, see Amazon EBS Encryption (p. 865).
10. (Optional) Choose Create additional tags to add tags to the volume. For each tag, provide a tag key and a tag value. For more information, see Tagging Your Amazon EC2 Resources (p. 964).
11. Choose Create Volume. After the volume status is Available, you can attach the volume to an instance. For more information, see Attaching an Amazon EBS Volume to an Instance (p. 803).

To create a new (empty) EBS volume using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
Restoring an Amazon EBS Volume from a Snapshot

You can restore an Amazon EBS volume with data from a snapshot stored in Amazon S3. You must know the ID of the snapshot and you must have access permissions for the snapshot. For more information about snapshots, see Amazon EBS Snapshots (p. 818).

EBS snapshots are the preferred backup tool on Amazon EC2 due to their speed, convenience, and cost. When restoring a volume from a snapshot, you recreate its state at a specific point in the past with all data intact. By attaching a restored volume to an instance, you can duplicate data across regions, create test environments, replace a damaged or corrupted production volume in its entirety, or retrieve specific files and directories and transfer them to another attached volume. For more information, see Amazon EBS Snapshots (p. 818).

New volumes created from existing EBS snapshots load lazily in the background. This means that after a volume is created from a snapshot, there is no need to wait for all of the data to transfer from Amazon S3 to your EBS volume before your attached instance can start accessing the volume and all its data. If your instance accesses data that hasn't yet been loaded, the volume immediately downloads the requested data from Amazon S3, and then continues loading the rest of the volume data in the background.

EBS Performance

New EBS volumes receive their maximum performance the moment that they are available and do not require initialization (formerly known as pre-warming). However, storage blocks on volumes that were restored from snapshots must be initialized (pulled down from Amazon S3 and written to the volume) before you can access the block. This preliminary action takes time and can cause a significant increase in the latency of an I/O operation the first time each block is accessed. Performance is restored after the data is accessed once.

For most applications, amortizing the initialization cost over the lifetime of the volume is acceptable. To ensure that your restored volume always functions at peak capacity in production, you can force the immediate initialization of the entire volume using `dd` or `fio`. For more information, see Initializing Amazon EBS Volumes (p. 891).

EBS Encryption

New EBS volumes that are restored from encrypted snapshots are automatically encrypted. You can also encrypt a volume on-the-fly while restoring it from an unencrypted snapshot. Encrypted volumes can only be attached to instance types that support EBS encryption. For more information, see Supported Instance Types (p. 866).

Because of security constraints, you cannot directly restore an EBS volume from a shared encrypted snapshot that you do not own. You must first create a copy of the snapshot, which you will own. You can then restore a volume from that copy. For more information, see Encryption and Snapshot Copying (p. 839).

Creating a Volume from a Snapshot

Use the following procedure to create a volume from a snapshot.

**To create an EBS volume from a snapshot using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the region that your snapshot is located in.
To restore the snapshot to a volume in a different region, you can copy your snapshot to the new region and then restore it to a volume in that region. For more information, see Copying an Amazon EBS Snapshot (p. 837).

3. In the navigation pane, choose ELASTIC BLOCK STORE, Volumes.
4. Choose Create Volume.
5. For Volume Type, choose a volume type. For more information, see Amazon EBS Volume Types (p. 786).
6. For Snapshot ID, start typing the ID or description of the snapshot from which you are restoring the volume, and choose it from the list of suggested options.
7. (Optional) Select Encrypt this volume to change the encryption state of your volume. This is optional if encryption by default (p. 868) is enabled. Select a CMK from Master Key to specify a CMK other than the default CMK for EBS encryption.
8. For Size (GiB), type the size of the volume, or verify that the default size of the snapshot is adequate.

If you specify both a volume size and a snapshot, the size must be equal to or greater than the snapshot size. When you select a volume type and a snapshot, the minimum and maximum sizes for the volume are shown next to Size. For more information, see Constraints on the Size and Configuration of an EBS Volume (p. 798).

9. With a Provisioned IOPS SSD volume, for IOPS, type the maximum number of input/output operations per second (IOPS) that the volume should support.
10. For Availability Zone, choose the Availability Zone in which to create the volume. EBS volumes can only be attached to EC2 instances in the same Availability Zone.
11. (Optional) Choose Create additional tags to add tags to the volume. For each tag, provide a tag key and a tag value.
12. Choose Create Volume.
13. After you've restored a volume from a snapshot, you can attach it to an instance to begin using it. For more information, see Attaching an Amazon EBS Volume to an Instance (p. 803).
14. If you restored a snapshot to a larger volume than the default for that snapshot, you must extend the file system on the volume to take advantage of the extra space. For more information, see Amazon EBS Elastic Volumes (p. 855).

To create an EBS volume from a snapshot using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- create-volume (AWS CLI)
- New-EC2Volume (AWS Tools for Windows PowerShell)

Attaching an Amazon EBS Volume to an Instance

You can attach an available EBS volume to one of your instances that is in the same Availability Zone as the volume.

Prerequisites

- Determine how many volumes you can attach to your instance. For more information, see Instance Volume Limits (p. 928).
- If a volume is encrypted, it can only be attached to an instance that supports Amazon EBS encryption. For more information, see Supported Instance Types (p. 866).
• If a volume has an AWS Marketplace product code:
  • The volume can only be attached to a stopped instance.
  • You must be subscribed to the AWS Marketplace code that is on the volume.
  • The configuration (instance type, operating system) of the instance must support that specific AWS
    Marketplace code. For example, you cannot take a volume from a Windows instance and attach it to
    a Linux instance.
  • AWS Marketplace product codes are copied from the volume to the instance.

**To attach an EBS volume to an instance using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose Elastic Block Store, Volumes.
3. Select an available volume and choose Actions, Attach Volume.
4. For Instance, start typing the name or ID of the instance. Select the instance from the list of options
   (only instances that are in the same Availability Zone as the volume are displayed).
5. For Device, you can keep the suggested device name, or type a different supported device name. For
   more information, see Device Naming on Windows Instances (p. 929).
6. Choose Attach.
7. Connect to your instance and mount the volume. For more information, see Making an Amazon EBS
   Volume Available for Use on Windows (p. 804).

**To attach an EBS volume to an instance using the command line**

You can use one of the following commands. For more information about these command line interfaces,
see Accessing Amazon EC2 (p. 3).

- attach-volume (AWS CLI)
- Add-EC2Volume (AWS Tools for Windows PowerShell)

**Making an Amazon EBS Volume Available for Use on Windows**

After you attach an Amazon EBS volume to your instance, it is exposed as a block device, and appears
as a removable disk in Windows. You can format the volume with any file system and then mount it.
After you make the EBS volume available for use, you can access it in the same ways that you access any
other volume. Any data written to this file system is written to the EBS volume and is transparent to
applications using the device.

You can take snapshots of your EBS volume for backup purposes or to use as a baseline when you create
another volume. For more information, see Amazon EBS Snapshots (p. 818).

You can get directions for volumes on a Linux instance from Making a Volume Available for Use on Linux
in the Amazon EC2 User Guide for Linux Instances.

You can make an EBS volume available for use using the Disk Management utility and the DiskPart
command line tool.

**To make an EBS volume available for use using the Disk Management utility**

1. Log in to your Windows instance using Remote Desktop. For more information, see, Connecting to
   Your Windows Instance (p. 379).
2. Start the Disk Management utility. On the taskbar, open the context (right-click) menu for the
   Windows logo and choose Disk Management.
Note

3. Bring the volume online. In the lower pane, open the context (right-click) menu for the left panel for the disk for the EBS volume. Choose Online.

4. (Conditional) You must initialize the disk before you can use it.

   Warning
   If you’re mounting a volume that already has data on it (for example, a public data set, or a volume that you created from a snapshot), do not reformat the volume or you will delete the existing data.

   If the disk is not initialized, initialize it as follows. Open the context (right-click) menu for the left panel for the disk and choose Initialize Disk. In the Initialize Disk dialog box, select a partition style and choose OK.
5. Open the context (right-click) menu for the right panel for the disk and choose New Simple Volume. Complete the wizard.
To make an EBS volume available for use using the DiskPart command line tool

1. Log in to your Windows instance using Remote Desktop. For more information, see Connecting to Your Windows Instance (p. 379).
2. Create a new script file named diskpart.txt.
3. Add the following commands to the script file and specify the volume label and drive letter. This script configures the volume to use the master boot record (MBR) partition structure, formats the volume as an NTFS volume, sets the volume label, and assigns it a drive letter.

   **Warning**
   If you’re mounting a volume that already has data on it, do not reformat the volume or you will delete the existing data.

```
select disk 1
attributes disk clear readonly
online disk
convert mbr
create partition primary
format quick fs=ntfs label="volume_label"
assign letter="drive_letter"
```

For more information, see DiskPart Syntax and Parameters.
4. Navigate to the folder in which the script is located and execute the following command:

```
C:\> diskpart /s diskpart.txt
```

Viewing Information about an Amazon EBS Volume

You can view descriptive information about your EBS volumes. For example, you can view information about all volumes in a specific region or view detailed information about a single volume, including its size, volume type, whether the volume is encrypted, which master key was used to encrypt the volume, and the specific instance to which the volume is attached.

You can get additional information about your EBS volumes, such as how much disk space is available, from the operating system on the instance.

Viewing Descriptive Information

To view information about an EBS volume using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Volumes.
3. To view more information about a volume, select it. In the details pane, you can inspect the information provided about the volume.
4. In the details pane, you can inspect the information provided about the volume.

To view the EBS volumes that are attached to an instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. To view more information about an instance, select it.
4. In the details pane, you can inspect the information provided about root and block devices.
To view information about an EBS volume using the command line

You can use one of the following commands to view volume attributes. For more information, see Accessing Amazon EC2 (p. 3).

- `describe-volumes` (AWS CLI)
- `Get-EC2Volume` (AWS Tools for Windows PowerShell)

Viewing Free Disk Space

You can get additional information about your EBS volumes, such as how much disk space is available, from the Windows operating system on the instance. For example, you can view the free disk space by opening File Explorer and selecting This PC.

You can also view the free disk space using the following `dir` command and examining the last line of the output:

```
C:\> dir C:
Volume in drive C has no label.
Volume Serial Number is 68C3-8081
Directory of C:\

03/25/2018  02:10 AM    <DIR>          .
03/25/2018  02:10 AM    <DIR>          ..
03/25/2018  03:47 AM    <DIR>          Contacts
03/25/2018  03:47 AM    <DIR>          Desktop
03/25/2018  03:47 AM    <DIR>          Documents
03/25/2018  03:47 AM    <DIR>          Downloads
03/25/2018  03:47 AM    <DIR>          Favorites
03/25/2018  03:47 AM    <DIR>          Links
03/25/2018  03:47 AM    <DIR>          Music
03/25/2018  03:47 AM    <DIR>          Pictures
03/25/2018  03:47 AM    <DIR>          Saved Games
03/25/2018  03:47 AM    <DIR>          Searches
03/25/2018  03:47 AM    <DIR>          Videos
0 File(s)              0 bytes
13 Dir(s) 13,163,839,936 bytes free
```

You can also view the free disk space using the following `fsutil` command:

```
C:\> fsutil volume diskfree C:
Total # of free bytes : 18113204224
Total # of bytes      : 32210153472
Total # of avail free bytes : 18113204224
```

Monitoring the Status of Your Volumes

Amazon Web Services (AWS) automatically provides data that you can use to monitor your Amazon Elastic Block Store (Amazon EBS) volumes.

Contents

- EBS Volume Status Checks (p. 809)
- EBS Volume Events (p. 811)
- Working with an Impaired Volume (p. 812)
- Working with the Auto-Enabled IO Volume Attribute (p. 815)
For additional monitoring information, see Amazon CloudWatch Metrics for Amazon EBS (p. 903) and Amazon CloudWatch Events for Amazon EBS (p. 907).

**EBS Volume Status Checks**

Volume status checks enable you to better understand, track, and manage potential inconsistencies in the data on an Amazon EBS volume. They are designed to provide you with the information that you need to determine whether your Amazon EBS volumes are impaired, and to help you control how a potentially inconsistent volume is handled.

Volume status checks are automated tests that run every 5 minutes and return a pass or fail status. If all checks pass, the status of the volume is `ok`. If a check fails, the status of the volume is `impaired`. If the status is `insufficient-data`, the checks may still be in progress on the volume. You can view the results of volume status checks to identify any impaired volumes and take any necessary actions.

When Amazon EBS determines that a volume's data is potentially inconsistent, the default is that it disables I/O to the volume from any attached EC2 instances, which helps to prevent data corruption. After I/O is disabled, the next volume status check fails, and the volume status is `impaired`. In addition, you'll see an event that lets you know that I/O is disabled, and that you can resolve the impaired status of the volume by enabling I/O to the volume. We wait until you enable I/O to give you the opportunity to decide whether to continue to let your instances use the volume, or to run a consistency check using a command, such as `chkdsk`, before doing so.

**Note**

Volume status is based on the volume status checks, and does not reflect the volume state. Therefore, volume status does not indicate volumes in the *error* state (for example, when a volume is incapable of accepting I/O.)

If the consistency of a particular volume is not a concern, and you'd prefer that the volume be made available immediately if it's impaired, you can override the default behavior by configuring the volume to automatically enable I/O. If you enable the `Auto-Enable IO` volume attribute (`autoEnableIO` in the API), the volume status check continues to pass. In addition, you'll see an event that lets you know that the volume was determined to be potentially inconsistent, but that its I/O was automatically enabled. This enables you to check the volume's consistency or replace it at a later time.

The I/O performance status check compares actual volume performance to the expected performance of a volume and alerts you if the volume is performing below expectations. This status check is only available for `io1` volumes that are attached to an instance and is not valid for General Purpose SSD (gp2), Throughput Optimized HDD (st1), Cold HDD (sc1), or Magnetic (standard) volumes. The I/O performance status check is performed once every minute and CloudWatch collects this data every 5 minutes, so it may take up to 5 minutes from the moment you attach a `io1` volume to an instance for this check to report the I/O performance status.

**Important**

While initializing `io1` volumes that were restored from snapshots, the performance of the volume may drop below 50 percent of its expected level, which causes the volume to display a `warning` state in the I/O Performance status check. This is expected, and you can ignore the warning state on `io1` volumes while you are initializing them. For more information, see Initializing Amazon EBS Volumes (p. 891).

The following table lists statuses for Amazon EBS volumes.

<table>
<thead>
<tr>
<th>Volume status</th>
<th>I/O enabled status</th>
<th>I/O performance status (only available for Provisioned IOPS volumes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ok</td>
<td>Enabled (I/O Enabled or I/O Auto-Enabled)</td>
<td>Normal (Volume performance is as expected)</td>
</tr>
</tbody>
</table>
### EBS Volumes

<table>
<thead>
<tr>
<th>Volume status</th>
<th>I/O enabled status</th>
<th>I/O performance status (only available for Provisioned IOPS volumes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>warning</td>
<td>Enabled (I/O Enabled or I/O Auto-Enabled)</td>
<td>Degraded (Volume performance is below expectations)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severely Degraded (Volume performance is well below expectations)</td>
</tr>
<tr>
<td>impaired</td>
<td>Enabled (I/O Enabled or I/O Auto-Enabled)</td>
<td>Stalled (Volume performance is severely impacted)</td>
</tr>
<tr>
<td></td>
<td>Disabled (Volume is offline and pending recovery, or is waiting for the user to enable I/O)</td>
<td>Not Available (Unable to determine I/O performance because I/O is disabled)</td>
</tr>
<tr>
<td>insufficient-data</td>
<td>Enabled (I/O Enabled or I/O Auto-Enabled)</td>
<td>Insufficient Data</td>
</tr>
<tr>
<td></td>
<td>Insufficient Data</td>
<td></td>
</tr>
</tbody>
</table>

To view and work with status checks, you can use the Amazon EC2 console, the API, or the command line interface.

**To view status checks in the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Volumes**. The **Volume Status** column displays the operational status of each volume.
3. To view the status details of a volume, select the volume and choose **Status Checks**.

4. If you have a volume with a failed status check (status is **impaired**), see **Working with an Impaired Volume (p. 812)**.
Alternatively, you can choose **Events** in the navigator to view all the events for your instances and volumes. For more information, see EBS Volume Events (p. 811).

**To view volume status information with the command line**

You can use one of the following commands to view the status of your Amazon EBS volumes. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `describe-volume-status` (AWS CLI)
- `Get-EC2VolumeStatus` (AWS Tools for Windows PowerShell)

**EBS Volume Events**

When Amazon EBS determines that a volume's data is potentially inconsistent, it disables I/O to the volume from any attached EC2 instances by default. This causes the volume status check to fail, and creates a volume status event that indicates the cause of the failure.

To automatically enable I/O on a volume with potential data inconsistencies, change the setting of the **Auto-Enabled IO** volume attribute (`autoEnableIO` in the API). For more information about changing this attribute, see Working with an Impaired Volume (p. 812).

Each event includes a start time that indicates the time at which the event occurred, and a duration that indicates how long I/O for the volume was disabled. The end time is added to the event when I/O for the volume is enabled.

Volume status events include one of the following descriptions:

**Awaiting Action: Enable IO**

- Volume data is potentially inconsistent. I/O is disabled for the volume until you explicitly enable it. The event description changes to **IO Enabled** after you explicitly enable I/O.

**IO Enabled**

- I/O operations were explicitly enabled for this volume.

**IO Auto-Enabled**

- I/O operations were automatically enabled on this volume after an event occurred. We recommend that you check for data inconsistencies before continuing to use the data.

**Normal**

- For io1 volumes only. Volume performance is as expected.

**Degraded**

- For io1 volumes only. Volume performance is below expectations.

**Severely Degraded**

- For io1 volumes only. Volume performance is well below expectations.

**Stalled**

- For io1 volumes only. Volume performance is severely impacted.

You can view events for your volumes using the Amazon EC2 console, the API, or the command line interface.

**To view events for your volumes in the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Events**. All instances and volumes that have events are listed.
3. You can filter by volume to view only volume status. You can also filter on specific status types.
4. Select a volume to view its specific event.

If you have a volume where I/O is disabled, see **Working with an Impaired Volume** (p. 812). If you have a volume where I/O performance is below normal, this might be a temporary condition due to an action you have taken (for example, creating a snapshot of a volume during peak usage, running the volume on an instance that cannot support the I/O bandwidth required, accessing data on the volume for the first time, etc.).

**To view events for your volumes with the command line**

You can use one of the following commands to view event information for your Amazon EBS volumes. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).

- `describe-volume-status` (AWS CLI)
- `Get-EC2VolumeStatus` (AWS Tools for Windows PowerShell)

**Working with an Impaired Volume**

Use the following options if a volume is impaired because the volume's data is potentially inconsistent.
Options
- Option 1: Perform a Consistency Check on the Volume Attached to its Instance (p. 813)
- Option 2: Perform a Consistency Check on the Volume Using Another Instance (p. 814)
- Option 3: Delete the Volume If You No Longer Need It (p. 815)

Option 1: Perform a Consistency Check on the Volume Attached to its Instance

The simplest option is to enable I/O and then perform a data consistency check on the volume while the volume is still attached to its Amazon EC2 instance.

To perform a consistency check on an attached volume

1. Stop any applications from using the volume.
2. Enable I/O on the volume.
   a. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
   b. In the navigation pane, choose Volumes.
   c. Select the volume on which to enable I/O operations.
   d. In the details pane, choose Enable Volume IO, and then choose Yes, Enable.

3. Check the data on the volume.
   a. Run the chkdsk command.
   b. (Optional) Review any available application or system logs for relevant error messages.
   c. If the volume has been impaired for more than 20 minutes, you can contact the AWS Support Center. Choose Troubleshoot, and then in the Troubleshoot Status Checks dialog box, choose Contact Support to submit a support case.

To enable I/O for a volume with the command line

You can use one of the following commands to view event information for your Amazon EBS volumes. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- enable-volume-io (AWS CLI)
- Enable-EC2VolumeIO (AWS Tools for Windows PowerShell)
Option 2: Perform a Consistency Check on the Volume Using Another Instance

Use the following procedure to check the volume outside your production environment.

**Important**
This procedure may cause the loss of write I/Os that were suspended when volume I/O was disabled.

**To perform a consistency check on a volume in isolation**

1. Stop any applications from using the volume.
2. Detach the volume from the instance.
   a. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
   b. In the navigation pane, choose **Volumes**.
   c. Select the volume to detach.
   d. Choose **Actions**, **Force Detach Volume**. You’ll be prompted for confirmation.
3. Enable I/O on the volume.
   a. In the navigation pane, choose **Volumes**.
   b. Select the volume that you detached in the previous step.
   c. In the details pane, choose **Enable Volume IO**, and then choose **Yes, Enable**.
4. Attach the volume to another instance. For more information, see Launch Your Instance (p. 333) and Attaching an Amazon EBS Volume to an Instance (p. 803).
5. Check the data on the volume.
   a. Run the `chkdsk` command.
   b. (Optional) Review any available application or system logs for relevant error messages.
   c. If the volume has been impaired for more than 20 minutes, you can contact the AWS Support Center. Choose **Troubleshoot**, and then in the troubleshooting dialog box, choose **Contact Support** to submit a support case.

**To enable I/O for a volume with the command line**

You can use one of the following commands to view event information for your Amazon EBS volumes. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).
• enable-volume-io (AWS CLI)
• Enable-EC2VolumeIO (AWS Tools for Windows PowerShell)

Option 3: Delete the Volume If You No Longer Need It

If you want to remove the volume from your environment, simply delete it. For information about deleting a volume, see Deleting an Amazon EBS Volume (p. 818).

If you have a recent snapshot that backs up the data on the volume, you can create a new volume from the snapshot. For information about creating a volume from a snapshot, see Restoring an Amazon EBS Volume from a Snapshot (p. 802).

Working with the Auto-Enabled IO Volume Attribute

When Amazon EBS determines that a volume's data is potentially inconsistent, it disables I/O to the volume from any attached EC2 instances by default. This causes the volume status check to fail, and creates a volume status event that indicates the cause of the failure. If the consistency of a particular volume is not a concern, and you prefer that the volume be made available immediately if it's impaired, you can override the default behavior by configuring the volume to automatically enable I/O. If you enable the Auto-Enabled IO volume attribute (autoEnableIO in the API), I/O between the volume and the instance is automatically re-enabled and the volume's status check will pass. In addition, you'll see an event that lets you know that the volume was in a potentially inconsistent state, but that its I/O was automatically enabled. When this event occurs, you should check the volume's consistency and replace it if necessary. For more information, see EBS Volume Events (p. 811).

This procedure explains how to view and modify the Auto-Enabled IO attribute of a volume.

To view the Auto-Enabled IO attribute of a volume in the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Volumes.
3. Select the volume and choose Status Checks. Auto-Enabled IO displays the current setting (Enabled or Disabled) for your volume.
To modify the Auto-Enabled IO attribute of a volume in the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Volumes.
3. Select the volume and choose Actions, Change Auto-Enable IO Setting. Alternatively, choose the Status Checks tab, and for Auto-Enabled IO, choose Edit.

4. Select the Auto-Enable Volume IO check box to automatically enable I/O for an impaired volume. To disable the feature, clear the check box.

5. Choose Save.

To view or modify the autoEnableIO attribute of a volume with the command line

You can use one of the following commands to view the autoEnableIO attribute of your Amazon EBS volumes. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- describe-volume-attribute (AWS CLI)
- Get-EC2VolumeAttribute (AWS Tools for Windows PowerShell)

To modify the autoEnableIO attribute of a volume, you can use one of the commands below.

- modify-volume-attribute (AWS CLI)
- Edit-EC2VolumeAttribute (AWS Tools for Windows PowerShell)

Detaching an Amazon EBS Volume from an Instance

You can detach an Amazon EBS volume from an instance explicitly or by terminating the instance. However, if the instance is running, you must first unmount the volume from the instance.
If an EBS volume is the root device of an instance, you must stop the instance before you can detach the volume.

When a volume with an AWS Marketplace product code is detached from an instance, the product code is no longer associated with the instance.

**Important**

After you detach a volume, you are still charged for volume storage as long as the storage amount exceeds the limit of the AWS Free Tier. You must delete a volume to avoid incurring further charges. For more information, see Deleting an Amazon EBS Volume (p. 818).

This example unmounts the volume and then explicitly detaches it from the instance. This is useful when you want to terminate an instance or attach a volume to a different instance. To verify that the volume is no longer attached to the instance, see Viewing Information about an Amazon EBS Volume (p. 807).

You can reattach a volume that you detached (without unmounting it), but it might not get the same mount point. If there were writes to the volume in progress when it was detached, the data on the volume might be out of sync.

### To detach an EBS volume using the console

1. Unmount the volume. Choose Disk Management, right-click the volume, and then choose Change Drive Letter and Path. Select the mount point and choose Remove.
2. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
3. In the navigation pane, choose Volumes.
4. Select a volume and choose Actions, Detach Volume.
5. In the confirmation dialog box, choose Yes, Detach.

### To detach an EBS volume from an instance using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `detach-volume` (AWS CLI)
- `Dismount-EC2Volume` (AWS Tools for Windows PowerShell)

### Troubleshooting

The following are common problems encountered when detaching volumes, and how to resolve them.

**Note**

To guard against the possibility of data loss, take a snapshot of your volume before attempting to unmount it. Forced detachment of a stuck volume can cause damage to the file system or the data it contains or an inability to attach a new volume using the same device name, unless you reboot the instance.

- If you encounter problems while detaching a volume through the Amazon EC2 console, it may be helpful to use the `describe-volumes` CLI command to diagnose the issue. For more information, see `describe-volumes`.
- If your volume stays in the `detaching` state, you can force the detachment by choosing Force Detach. Use this option only as a last resort to detach a volume from a failed instance, or if you are detaching a volume with the intention of deleting it. The instance doesn't get an opportunity to flush file system caches or file system metadata. If you use this option, you must perform the file system check and repair procedures.
- If you've tried to force the volume to detach multiple times over several minutes and it stays in the `detaching` state, you can post a request for help to the Amazon EC2 forum. To help expedite a resolution, include the volume ID and describe the steps that you've already taken.
When you attempt to detach a volume that is still mounted, the volume can become stuck in the busy state while it is trying to detach. The following output from describe-volumes shows an example of this condition:

```
aws ec2 describe-volumes --region us-west-2 --volume-ids vol-1234abcd
{
  "Volumes": [
    {
      "AvailabilityZone": "us-west-2b",
      "Attachments": [
        {
          "AttachTime": "2016-07-21T23:44:52.000Z",
          "InstanceId": "i-fedc9876",
          "VolumeId": "vol-1234abcd",
          "State": "busy",
          "DeleteOnTermination": false,
          "Device": "/dev/sdf"
        }
      ...
    }
  }
}
```

When you encounter this state, detachment can be delayed indefinitely until you unmount the volume, force detachment, reboot the instance, or all three.

**Deleting an Amazon EBS Volume**

After you no longer need an Amazon EBS volume, you can delete it. After deletion, its data is gone and the volume can't be attached to any instance. However, before deletion, you can store a snapshot of the volume, which you can use to re-create the volume later.

To delete a volume, it must be in the available state (not attached to an instance). For more information, see Detaching an Amazon EBS Volume from an Instance (p. 816).

**To delete an EBS volume using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Volumes**.
3. Select a volume and choose **Actions, Delete Volume**.
4. In the confirmation dialog box, choose **Yes, Delete**.

**To delete an EBS volume using the command line**

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `delete-volume` (AWS CLI)
- `Remove-EC2Volume` (AWS Tools for Windows PowerShell)

**Amazon EBS Snapshots**

You can back up the data on your Amazon EBS volumes to Amazon S3 by taking point-in-time snapshots. Snapshots are incremental backups, which means that only the blocks on the device that have changed after your most recent snapshot are saved. This minimizes the time required to create the snapshot and saves on storage costs by not duplicating data. When you delete a snapshot, only the data unique to that snapshot is removed. Each snapshot contains all of the information that is needed to restore your data (from the moment when the snapshot was taken) to a new EBS volume.
When you create an EBS volume based on a snapshot, the new volume begins as an exact replica of the original volume that was used to create the snapshot. The replicated volume loads data in the background so that you can begin using it immediately. If you access data that hasn't been loaded yet, the volume immediately downloads the requested data from Amazon S3, and then continues loading the rest of the volume's data in the background. For more information, see Creating Amazon EBS Snapshots (p. 822).

**Note**
Using Systems Manager Run Command, you can take application-consistent snapshots of all EBS volumes attached to your Amazon EC2 Windows instances. The snapshot process uses the Windows Volume Shadow Copy Service (VSS) to take image-level backups of VSS-aware applications, including data from pending transactions between these applications and the disk. You don't need to shut down your instances or disconnect them when you back up all attached volumes. For more information, see Using Run Command to Take VSS-Enabled Snapshots of EBS Volumes in the AWS Systems Manager User Guide.

**Multi-Volume Snapshots**
Snapshots can be used to create a backup of critical workloads, such as a large database or a file system that spans across multiple EBS volumes. Multi-volume snapshots allow you to take exact point-in-time, data coordinated, and crash-consistent snapshots across multiple EBS volumes attached to an EC2 instance. You are no longer required to stop your instance or to coordinate between volumes to ensure crash consistency, because snapshots are automatically taken across multiple EBS volumes. For more information, see the steps for creating a multi-volume EBS snapshot under Creating Amazon EBS Snapshots (p. 822).

You can track the status of your EBS snapshots through CloudWatch Events. For more information, see Amazon CloudWatch Events for Amazon EBS (p. 907).

**Contents**
- How Incremental Snapshots Work (p. 819)
- Copying and Sharing Snapshots (p. 822)
- Encryption Support for Snapshots (p. 822)
- Creating Amazon EBS Snapshots (p. 822)
- Creating a VSS Application-Consistent Snapshot (p. 825)
- Deleting an Amazon EBS Snapshot (p. 835)
- Copying an Amazon EBS Snapshot (p. 837)
- Viewing Amazon EBS Snapshot Information (p. 841)
- Sharing an Amazon EBS Snapshot (p. 842)
- Automating the Amazon EBS Snapshot Lifecycle (p. 844)

**How Incremental Snapshots Work**
This section provides illustrations of how an EBS snapshot captures the state of a volume at a point in time, and also how successive snapshots of a changing volume create a history of those changes.

In the diagram below, Volume 1 is shown at three points in time. A snapshot is taken of each of these three volume states:

- In State 1, the volume has 10 GiB of data. Because Snap A is the first snapshot taken of the volume, the entire 10 GiB of data must be copied.
- In State 2, the volume still contains 10 GiB of data, but 4 GiB have changed. Snap B needs to copy and store only the 4 GiB that changed after Snap A was taken. The other 6 GiB of unchanged data, which are already copied and stored in Snap A, are referenced by Snap B rather than (again) copied. This is indicated by the dashed arrow.
• In State 3, 2 GiB of data have been added to the volume, for a total of 12 GiB. Snap C needs to copy the 2 GiB that were added after Snap B was taken. As shown by the dashed arrows, Snap C also references 4 GiB of data stored in Snap B, and 6 GiB of data stored in Snap A.

• The total storage required for the three snapshots is 16 GiB.

Relations among Multiple Snapshots of a Volume
Note
If you copy a snapshot and encrypt it to a new CMK, a complete (non-incremental) copy is always created, resulting in additional delay and storage costs.
Copy snapshots across AWS accounts by modifying access permissions. You can make copies of your own snapshots as well as snapshots that have been shared with you. For more information, see Sharing an Amazon EBS Snapshot (p. 842).

A snapshot is constrained to the AWS Region where it was created. After you create a snapshot of an EBS volume, you can use it to create new volumes in the same Region. For more information, see Restoring an Amazon EBS Volume from a Snapshot (p. 802). You can also copy snapshots across Regions, making it possible to use multiple Regions for geographical expansion, data center migration, and disaster recovery. You can copy any accessible snapshot that has a completed status. For more information, see Copying an Amazon EBS Snapshot (p. 837).

Encryption Support for Snapshots

EBS snapshots fully support EBS encryption.

- Snapshots of encrypted volumes are automatically encrypted.
- Volumes that you create from encrypted snapshots are automatically encrypted.
- Volumes that you create from an unencrypted snapshot that you own or have access to can be encrypted on-the-fly.
- When you copy an unencrypted snapshot that you own, you can encrypt it during the copy process.
- When you copy an encrypted snapshot that you own or have access to, you can reencrypt it with a different key during the copy process.
- The first snapshot you take of an encrypted volume that has been created from an unencrypted snapshot is always a full snapshot.
- The first snapshot you take of a reencrypted volume, which has a different CMK compared to the source snapshot, is always a full snapshot.

**Note**
If you copy a snapshot and encrypt it to a new CMK, a complete (non-incremental) copy is always created, resulting in additional delay and storage costs.

Complete documentation of possible snapshot encryption scenarios is provided in Creating Amazon EBS Snapshots (p. 822) and in Copying an Amazon EBS Snapshot (p. 837).

For more information, see Amazon EBS Encryption (p. 865).

Creating Amazon EBS Snapshots

To create an application-consistent snapshot, see Creating a VSS Application-Consistent Snapshot (p. 825).

You can create a point-in-time snapshot of an EBS volume and use it as a baseline for new volumes or for data backup. If you make periodic snapshots of a volume, the snapshots are incremental—the new snapshot saves only the blocks that have changed since your last snapshot.

Snapshots occur asynchronously; the point-in-time snapshot is created immediately, but the status of the snapshot is pending until the snapshot is complete (when all of the modified blocks have been transferred to Amazon S3), which can take several hours for large initial snapshots or subsequent
snapshots where many blocks have changed. While it is completing, an in-progress snapshot is not
affected by ongoing reads and writes to the volume.

You can take a snapshot of an attached volume that is in use. However, snapshots only capture data
that has been written to your Amazon EBS volume at the time the snapshot command is issued. This
might exclude any data that has been cached by any applications or the operating system. If you can
pause any file writes to the volume long enough to take a snapshot, your snapshot should be complete.
However, if you can't pause all file writes to the volume, you should unmount the volume from within
the instance, issue the snapshot command, and then remount the volume to ensure a consistent and
complete snapshot. You can remount and use your volume while the snapshot status is pending.

To make snapshot management easier, you can tag your snapshots during creation or add tags
afterward. For example, you can apply tags describing the original volume from which the snapshot
was created, or the device name that was used to attach the original volume to an instance. For more
information, see Tagging Your Amazon EC2 Resources (p. 964).

**Snapshot Encryption**

Snapshots that are taken from encrypted volumes are automatically encrypted. Volumes that are created
from encrypted snapshots are also automatically encrypted. The data in your encrypted volumes and
any associated snapshots is protected both at rest and in motion. For more information, see Amazon EBS
Encryption (p. 865).

By default, only you can create volumes from snapshots that you own. However, you can share
your unencrypted snapshots with specific AWS accounts, or you can share them with the entire
AWS community by making them public. For more information, see Sharing an Amazon EBS
Snapshot (p. 842).

You can share an encrypted snapshot only with specific AWS accounts. For others to use your shared,
encrypted snapshot, you must also share the CMK key that was used to encrypt it. Users with access to
your encrypted snapshot must create their own personal copy of it and then use that copy to restore the
volume. Your copy of a shared, encrypted snapshot can also be re-encrypted with a different key. For
more information, see Sharing an Amazon EBS Snapshot (p. 842).

**Note**

If you copy a snapshot and encrypt it to a new CMK, a complete (non-incremental) copy is
always created, resulting in additional delay and storage costs.

**Multi-Volume Snapshots**

You can create multi-volume snapshots, which are point-in-time snapshots for all EBS volumes
attached to a single EC2 instance. You can also create lifecycle policies to automate the creation and
retention of multi-volume snapshots. For more information, see Automating the Amazon EBS Snapshot
Lifecycle (p. 844).

After the snapshots are created, each snapshot is treated as an individual snapshot. You can perform
all snapshot operations, such as restore, delete, and cross-region/account copy, just as you would with
a single volume snapshot. You can also tag your multi-volume snapshots as you would a single volume
snapshot. We recommend you tag your multiple volume snapshots to manage them collectively during
restore, copy, or retention.

Multi-volume, crash-consistent snapshots are typically restored as a set. It is helpful to identify
the snapshots that are in a crash-consistent set by tagging your set with the instance ID, name, or
other relevant details. You can also choose to automatically copy tags from the source volume to
the corresponding snapshots. This helps you to set the snapshot metadata, such as access policies,
attachment information, and cost allocation, to match the source volume.

After it's created, a multi-volume snapshot behaves like any other snapshot. You can perform all
operations, such as restore, delete, and copy across Regions and accounts. You can also tag your
snapshots. We recommend that you tag your multi-volume snapshots to collectively manage them during restore, copy, or retention.

After creating your snapshots, they appear in your EC2 console created at the exact point-in-time. The snapshots are collectively managed and, therefore, if any one snapshot for the volume set fails, all of the other snapshots display an error status.

Considerations

The following considerations apply to creating snapshots:

- When you create a snapshot for an EBS volume that serves as a root device, you should stop the instance before taking the snapshot.
- You cannot create snapshots from instances for which hibernation is enabled.
- You cannot create snapshots from hibernated instances.
- Although you can take a snapshot of a volume while a previous snapshot of that volume is in the pending status, having multiple pending snapshots of a volume can result in reduced volume performance until the snapshots complete.
- There is a limit of five pending snapshots for a single gp2, io1, or Magnetic volume, and one pending snapshot for a single st1 or sc1 volume. If you receive a ConcurrentSnapshotLimitExceeded error while trying to create multiple concurrent snapshots of the same volume, wait for one or more of the pending snapshots to complete before creating another snapshot of that volume.
- When a snapshot is created from a volume with an AWS Marketplace product code, the product code is propagated to the snapshot.

Creating a Snapshot

Use the following procedure to create a snapshot from the specified volume.

To create a snapshot using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Snapshots under Elastic Block Store in the navigation pane.
3. Choose Create Snapshot.
4. For Select resource type, choose Volume.
5. For Volume, select the volume.
6. (Optional) Enter a description for the snapshot.
7. (Optional) Choose Add Tag to add tags to your snapshot. For each tag, provide a tag key and a tag value.
8. Choose Create Snapshot.

To create a snapshot using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- create-snapshot (AWS CLI)
- New-EC2Snapshot (AWS Tools for Windows PowerShell)

Creating a Multi-Volume Snapshot

Use the following procedure to create a snapshot from the volumes of an instance.
To create multi-volume snapshots using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Snapshots under Elastic Block Store in the navigation pane.
3. Choose Create Snapshot.
4. For Select resource type, choose Instance.
5. Select the instance ID for which you want to create simultaneous backups for all of the attached EBS volumes. Multi-volume snapshots support up to 40 EBS volumes per instance.
6. (Optional) Set Exclude root volume.
7. (Optional) Set Copy tags from volume flag to automatically copy tags from the source volume to the corresponding snapshots. This sets snapshot metadata—such as access policies, attachment information, and cost allocation—to match the source volume.
8. (Optional) Choose Add Tag to add tags to your snapshot. For each tag, provide a tag key and a tag value.
9. Choose Create Snapshot.

During snapshot creation, the snapshots are managed together. If one of the snapshots in the volume set fails, the other snapshots are moved to error status for the volume set. You can monitor the progress of your snapshots using CloudWatch Events. After the snapshot creation process completes, CloudWatch generates an event that contains the status and all of the relevant snapshots details for the affected instance.

To create multi-volume snapshots using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- create-snapshots (AWS CLI)
- New-EC2SnapshotBatch (AWS Tools for Windows PowerShell)

To create application-consistent snapshots using Systems Manager Run Command

You can use Systems Manager Run Command to take application-consistent snapshots of all EBS volumes attached to your Amazon EC2 Windows instances. The snapshot process uses the Windows Volume Shadow Copy Service (VSS) to take image-level backups of VSS-aware applications, including data from pending transactions between these applications and the disk. You don't need to shut down your instances or disconnect them when you back up all attached volumes. For more information, see Creating a VSS Application-Consistent Snapshot (p. 825).

Creating a VSS Application-Consistent Snapshot

You can take application-consistent snapshots of all Amazon Elastic Block Store (Amazon EBS) volumes attached to your Windows on Amazon EC2 instances by using AWS Systems Manager Run Command. The snapshot process uses the Windows Volume Shadow Copy Service (VSS) to take image-level backups of VSS-aware applications. The snapshots include data from pending transactions between these applications and the disk. You don't have to shut down your instances or disconnect them when you need to back up all attached volumes.

There is no additional cost to use VSS-enabled EBS snapshots. You only pay for EBS snapshots created by the backup process. For more information, see How is my EBS snapshot bill calculated?

Topics

- How It Works (p. 826)
How It Works

The process for taking application-consistent, VSS-enabled EBS snapshots consists of the following steps.

1. Complete Systems Manager prerequisites.
2. Enter parameters for the AWSEC2-CreateVssSnapshot SSM document and run this document by using Run Command. You can't create a VSS-enabled EBS snapshot for a specific volume. You can, however, specify a parameter to exclude the boot volume from the backup process.
3. The VSS agent on your instance coordinates all ongoing I/O operations for running applications.
4. The system flushes all I/O buffers and temporarily pauses all I/O operations. The pause lasts, at most, ten seconds.
5. During the pause, the system creates snapshots of all volumes attached to the instance.
6. The pause is lifted and I/O resumes operation.
7. The system adds all newly-created snapshots to the list of EBS snapshots. The system tags all VSS-enabled EBS snapshots successfully created by this process with `AppConsistent:true`. This tag helps you identify snapshots created by this process, as opposed to other processes. If the system encounters an error, the snapshot created by this process does not include the `AppConsistent:true` tag.
8. If you need to restore from a snapshot, you can use the standard EBS process of creating a volume from a snapshot, or you can restore all volumes to an instance by using a sample script, which is described later in this section.

Before You Begin

Before you create VSS-enabled EBS snapshots by using Run Command, review the following requirements and limitations, and complete the required tasks.

Amazon EC2 Windows instance requirements

VSS-enabled EBS snapshots are supported for instances running Windows Server 2008 R2 or later. (Windows Server 2008 R2 Core is currently not supported.) Verify that your instances meet all requirements for Amazon EC2 Windows. For more information, see Setting Up AWS Systems Manager in the AWS Systems Manager User Guide.

SSM Agent version

Update your instances to use SSM Agent version 2.2.58.0 or later. If you are using an older version of SSM Agent, you can update it by using Run Command. For more information, see Update SSM Agent by using Run Command in the AWS Systems Manager User Guide.

AWS Tools for Windows PowerShell version

Ensure that your instance is running version 3.3.48.0 or later of the AWS Tools for Windows PowerShell. To check your version number, run the following command on the instance:

```
Get-AWSPowerShellVersion
```

Getting Started

These instructions describe how to install the VSS components and perform an application-consistent snapshot of the EBS volumes attached to an EC2 Windows instance. For more information, see Getting Started with Amazon EC2 Windows Instances.

Contents

- Create an IAM Role for VSS-Enabled Snapshots (p. 827)
- Download and Install VSS Components to the Windows on EC2 Instance (p. 828)
- Creating a VSS Application-Consistent Snapshot Using the Console (p. 829)

Create an IAM Role for VSS-Enabled Snapshots

The following procedures describes how to work with IAM policies and IAM roles. The policy enables Systems Manager to create snapshots, tags snapshots, and attach metadata like a device ID to the default snapshot tags that the system creates.

To create an IAM policy for VSS-enabled snapshots

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Policies, and then choose Create policy.
3. On the Create policy page, choose the JSON tab, and then replace the default content with the following JSON policy.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "ec2:CreateTags",
            "Resource": "arn:aws:ec2:*::snapshot/*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeInstances",
                "ec2:CreateSnapshot"
            ],
            "Resource": "*"
        }
    ]
}
```

5. For Name, enter a name to identify the policy, such as VssSnapshotRole or another name that you prefer.
6. (Optional) For Description, enter a description of the role's purpose.
7. Choose Create policy.

Use the following procedure to create an IAM role for VSS-enabled snapshots. This role includes policies for Amazon EC2 and Systems Manager.
To create an IAM role for VSS-enabled EBS snapshots

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Roles, and then choose Create role.
3. Under Select type of trusted entity, choose AWS Service.
4. Immediately under Choose the service that will use this role, choose EC2, and then choose Next: Permissions.
5. Under Select your use case, choose EC2, and then choose Next: Permissions.
6. In the list of policies, choose the box next to AmazonSSMManagedInstanceCore. (Type SS in the search box if you need to narrow the list.)
7. Choose Next: Tags.
8. (Optional) Add one or more tag key-value pairs to organize, track, or control access for this role, and then choose Next: Review.
9. For Role name, enter a name for the role, such as VssSnapshotRole or another name that you prefer.
10. (Optional) For Role description, replace the default text with a description of this role's purpose.
11. Choose Create role. The system returns you to the Roles page.
12. Choose the role that you just created. The role Summary page opens.
13. Choose Attach policies.
14. Search for and choose the box next to the policy your created in the previous procedure, such as VssSnapshotRole or another name that you chose.
15. Choose Attach policy.
16. Attach this role to the instances for which you want to create VSS-enabled EBS snapshots. For more information, see Attaching an IAM Role to an Instance in the Amazon EC2 User Guide.

Download and Install VSS Components to the Windows on EC2 Instance

Systems Manager requires VSS components to be installed on your instances. Use the following procedure to install the components using the AWSVssComponents package. The package installs two components: a VSS requestor and a VSS provider.

2. In the navigation pane, choose Run Command.
3. Choose Run command.
4. For Command document, choose the button next to AWS-ConfigureAWSPackage.
5. For Command parameters, do the following:
   a. Verify that Action is set to Install.
   b. For Name, enter AWSVssComponents.
   c. For Version, leave the field empty so that Systems Manager installs the latest version.
6. For Targets, identify the instances on which you want to run this operation by specifying tags or selecting instances manually.

   **Note**
   If you choose to select instances manually, and an instance you expect to see is not included in the list, see Where Are My Instances? in the AWS Systems Manager User Guide for troubleshooting tips.
7. For Other parameters:
   a. (Optional) For Comment, type information about this command.
• For **Timeout (seconds)**, specify the number of seconds for the system to wait before failing the overall command execution.

8. (Optional) For **Rate control**:

• For **Concurrency**, specify either a number or a percentage of instances on which to run the command at the same time.

  **Note**
  If you selected targets by choosing Amazon EC2 tags, and you are not certain how many instances use the selected tags, then limit the number of instances that can run the document at the same time by specifying a percentage.

• For **Error threshold**, specify when to stop running the command on other instances after it fails on either a number or a percentage of instances. For example, if you specify three errors, then Systems Manager stops sending the command when the fourth error is received. Instances still processing the command might also send errors.

9. (Optional) For **Output options** section, if you want to save the command output to a file, select the box next to **Enable writing to an S3 bucket**. Specify the bucket and (optional) prefix (folder) names.

  **Note**
  The S3 permissions that grant the ability to write the data to an S3 bucket are those of the instance profile assigned to the instance, not those of the IAM user performing this task. For more information, see Create an IAM Instance Profile for Systems Manager in the *AWS Systems Manager User Guide*.

10. (Optional) Specify options for **SNS notifications**.

    For information about configuring Amazon SNS notifications for Run Command, see Configuring Amazon SNS Notifications for AWS Systems Manager.

11. Choose **Run**.

---

**Creating a VSS Application-Consistent Snapshot Using the Console**

Use the following procedure to create a VSS-enabled EBS snapshot.

**To create VSS-enabled EBS snapshots using the console**

2. In the navigation pane, choose **Run Command**.
3. Choose **Run command**.
4. For **Command document**, choose **AWSEC2-CreateVssSnapshot**.
5. For **Targets**, identify the instances on which you want to run this operation by specifying tags or selecting instances manually.

  **Note**
  If you choose to select instances manually, and an instance you expect to see is not included in the list, see Where Are My Instances? for troubleshooting tips.

6. For **Command parameters**, do the following:

   a. Choose an option from the **Exclude Boot Volume** list. Use this parameter to exclude boot volumes from the backup process.
   b. (Optional) For **Description** field, type a description. This description is applied to any snapshot created by this process.
   c. (Optional) For **Tags**, type keys and values for tags that you want to apply to any snapshot created by this process. Tags can help you locate, manage, and restore volumes from a list of snapshots. By default, the system populates the tag parameter with a Name key. For the value of this key, specify a name that you want to apply to snapshots created by this process.
If you want to specify additional tags, separate tags by using a semicolon. For example, Key=Environment,Value=Test;Key=User,Value=TestUser1.

We recommended that you tag snapshots. By default, the systems tags snapshots with the device ID, and AppConsistent (for indicating successful, application-consistent VSS-enabled EBS snapshots).

d. For Copy Only, choose True to perform a copy only backup operation. This option is set to False by default so that the system performs a full backup operation. A full backup operation prevents the system from breaking the differential backup chain in SQL Server when performing a backup.

Note
This option requires that AWS VSS provider version 1.2.00 or later be installed.

e. For No Writers, choose True to exclude application VSS writers from the snapshot process. This can help you resolve conflicts with third-party VSS backup components. This option is set to False by default.

Note
This option requires that AWS VSS provider version 1.2.00 or later be installed.

7. For Other parameters:
   - For Comment, type information about this command.
   - For Timeout (seconds), specify the number of seconds for the system to wait before failing the overall command execution.

8. (Optional) For Rate control:
   - For Concurrency, specify either a number or a percentage of instances on which to run the command at the same time.

   Note
   If you selected targets by choosing Amazon EC2 tags, and you are not certain how many instances use the selected tags, then limit the number of instances that can run the document at the same time by specifying a percentage.
   - For Error threshold, specify when to stop running the command on other instances after it fails on either a number or a percentage of instances. For example, if you specify three errors, then Systems Manager stops sending the command when the fourth error is received. Instances still processing the command might also send errors.

9. (Optional) For Output options, to save the command output to a file, select the box next to Enable writing to an S3 bucket. Specify the bucket and (optional) prefix (folder) names.

   Note
   The S3 permissions that grant the ability to write the data to an S3 bucket are those of the instance profile assigned to the instance, not those of the IAM user performing this task. For more information, see Setting Up Systems Manager.

10. (Optional) Specify options for SNS notifications.

   For information about configuring Amazon SNS notifications for Run Command, see Configuring Amazon SNS Notifications for AWS Systems Manager in the AWS Systems Manager User Guide.


If successful, the command populates the list of EBS snapshots with the new snapshots. You can locate these snapshots in the list of EBS snapshots by searching for the tags you specified, or by searching for AppConsistent. If the command execution failed, view the Systems Manager command output for details about why the execution failed. If the command successfully completed, but a specific volume backup failed, you can troubleshoot the failure in the list of EBS volumes.

If the command failed and you are using Systems Manager with VPC endpoints, verify that you configured the com.amazonaws.region.ec2 endpoint. Without the EC2 endpoint defined, the call
EBS Snapshots

to enumerate attached EBS volumes fails, which causes the Systems Manager command to fail. For more information about setting up VPC endpoints with Systems Manager, see Create a Virtual Private Cloud Endpoint in the AWS Systems Manager User Guide.

Note
You can automate backups by creating a maintenance window task that uses the AWSEC2-CreateVssSnapshot SSM document. For more information, see Working with Maintenance Windows (Console).

Creating a VSS Application-Consistent Snapshot Using the AWS CLI, AWS Tools for Windows PowerShell, or the AWSEC2-ManageVssIO SSM Document

This section includes procedures for creating VSS-enabled EBS snapshots by using the AWS CLI or AWS Tools for Windows PowerShell. It also contains an advanced method for creating VSS-enabled snapshots using the AWSEC2-ManageVssIO SSM document.

Topics
- Install the VSS Package Using the AWS CLI or Tools for Windows PowerShell (p. 831)
- Create VSS-Enabled EBS Snapshots Using the AWS CLI, Tools for Windows PowerShell, or the AWSEC2-ManageVssIO SSM Document (p. 832)

Install the VSS Package Using the AWS CLI or Tools for Windows PowerShell

Use one of the following command-line procedures to download and install the VSS components to the Windows on EC2 instance.

Install the VSS Package by Using the AWS CLI

Use the following procedure to download and install the AwsVssComponents package on your instances by using Run Command from the AWS CLI. The package installs two components: a VSS requestor and a VSS provider. The system copies these components to a directory on the instance, and then registers the provider DLL as a VSS provider.

To install the VSS package by using the AWS CLI

1. Install and configure the AWS CLI, if you have not already.
   For information, see Install or Upgrade and then Configure the AWS CLI in the AWS Systems Manager User Guide.
2. Run the following command to download and install the required VSS components for Systems Manager.

   aws ssm send-command --document-name "AWS-ConfigureAWSPackage" --instance-ids "i-12345678" --parameters '{"action": ["Install"], "name": ["AwsVssComponents"]}'

Install the VSS Package by Using Tools for Windows PowerShell

Use the following procedure to download and install the AwsVssComponents package on your instances by using Run Command from the Tools for Windows PowerShell. The package installs two components: a VSS requestor and a VSS provider. The system copies these components to a directory on the instance, and then registers the provider DLL as a VSS provider.

To install the VSS package by using AWS Tools for Windows PowerShell

1. Open AWS Tools for Windows PowerShell and run the following command to specify your credentials. You must either have administrator privileges in Amazon EC2 or have been granted the
appropriate permission in IAM. For more information, see Setting Up AWS Systems Manager in the AWS Systems Manager User Guide.

```
Set-AWSCredentials -AccessKey {key_name} -SecretKey {key_name}
```

2. Run the following command to set the Region for your PowerShell session. The example uses the us-east-2 Region.

```
Set-DefaultAWSRegion -Region us-east-2
```

3. Run the following command to download and install the required VSS components for Systems Manager.

```
Send-SSMCommand -DocumentName AWS-ConfigureAWSPackage -InstanceId "$instance" -Parameter @
{
'action'='Install'; 'name'='AwsVssComponents'
}
```

Create VSS-Enabled EBS Snapshots Using the AWS CLI, Tools for Windows PowerShell, or the AWSEC2-ManageVssIO SSM Document

Use one of the following command-line procedures to create VSS-enabled EBS snapshots.

Creating VSS-Enabled EBS Snapshots Using the AWS CLI

Use the following procedure to create VSS-enabled EBS snapshots by using the AWS CLI. When you run the command, you can specify the following parameters:

- Instance (Required): Specify one or more Amazon EC2 Windows instances. You can either manually specify instances, or you can specify tags.
- Description (Optional): Specify details about this backup.
- Tags (Optional): Specify key-value tag pairs that you want to assign to the snapshots. Tags can help you locate, manage, and restore volumes from a list of snapshots. By default, the system populates the tag parameter with a name key. For the value of this key, specify a name that you want to apply to snapshots created by this process. You can also add custom tags to this list by using the following format: `Key=Environment,Value=Test;Key=VssValue,Value=TestUser1`.

This parameter is optional, but we recommended that you tag snapshots. By default, the systems tags snapshots with the device ID, and `AppConsistent` (for indicating successful, application-consistent VSS-enabled EBS snapshots).
- Exclude Boot Volume (Optional): Use this parameter to exclude boot volumes from the backup process.

To create VSS-enabled EBS snapshots by using the AWS CLI

1. Install and configure the AWS CLI, if you have not already.

   For information, see Install or Upgrade and then Configure the AWS CLI in the AWS Systems Manager User Guide.

2. Run the following command to create VSS-enabled EBS snapshots.

```
aws ssm send-command --document-name "AWSEC2-CreateVssSnapshot" --instance-ids "i-12345678" --parameters "{"ExcludeBootVolume":false,"description":null,"Description"},"tags":null"
```

If successful, the command populates the list of EBS snapshots with the new snapshots. You can locate these snapshots in the list of EBS snapshots by searching for the tags you specified, or by searching for
Creating VSS-Enabled EBS Snapshots Using AWS Tools for Windows PowerShell

Use the following procedure to create VSS-enabled EBS snapshots by using the AWS Tools for Windows PowerShell. When you run the command, you can specify the following parameters:

- **Instance (Required):** Specify one or more Amazon EC2 Windows instances. You can either manually specify instances, or you can specify tags.
- **Description (Optional):** Specify details about this backup.
- **Tags (Optional):** Specify key-value tag pairs that you want to assign to the snapshots. Tags can help you locate, manage, and restore volumes from a list of snapshots. By default, the system populates the tag parameter with a name key. For the value of this key, specify a name that you want to apply to snapshots created by this process. You can also add custom tags to this list by using the following format: `Key=Environment,Value=Test;Key=User,Value=TestUser1`.

This parameter is optional, but we recommend that you tag snapshots. By default, the systems tags snapshots with the device ID, and AppConsistent (for indicating successful, application-consistent VSS-enabled EBS snapshots).
- **Exclude Boot Volume (Optional):** Use this parameter to exclude boot volumes from the backup process.

To create VSS-enabled EBS snapshots by using AWS Tools for Windows PowerShell

1. Open AWS Tools for Windows PowerShell and run the following command to specify your credentials. You must either have administrator privileges in Amazon EC2, or you must have been granted the appropriate permission in IAM. For more information, see Setting Up AWS Systems Manager in the AWS Systems Manager User Guide.

   ```powershell
   Set-AWSCredentials –AccessKey key_name –SecretKey key_name
   ```

2. Execute the following command to set the Region for your PowerShell session. The example uses the us-east-2 Region.

   ```powershell
   Set-DefaultAWSRegion -Region us-east-2
   ```

3. Execute the following command to create VSS-enabled EBS snapshots.

   ```powershell
   Send-SSMCommand -DocumentName AWSEC2-CreateVssSnapshot -InstanceId "$instance" -Parameter @{'ExcludeBootVolume'='False';'description'='a_description';'tags'='Key=key_name,Value=tag_value'}
   ```

   If successful, the command populates the list of EBS snapshots with the new snapshots. You can locate these snapshots in the list of EBS snapshots by searching for the tags you specified, or by searching for AppConsistent. If the command execution failed, view the command output for details about why the execution failed. If the command successfully completed, but a specific volume backup failed, you can troubleshoot the failure in the list of EBS snapshots.

   You can automate backups by creating a maintenance window task that uses the AWSEC2-CreateVssSnapshot SSM document. For more information, see Working with Maintenance Windows (Console) in the AWS Systems Manager User Guide.
Creating VSS-Enabled EBS Snapshots by Using the AWSEC2-ManageVssIO SSM Document (Advanced)

You can use the following script and the pre-defined AWSEC2-ManageVssIO SSM document to temporarily pause I/O, create VSS-enabled EBS snapshots, and restart I/O. This process runs in the context of the user who runs the command. If the user has sufficient permission to create and tag snapshots, then AWS Systems Manager can create and tag VSS-enabled EBS snapshots without the need for the additional IAM snapshot role on the instance.

In contrast, the AWSEC2-CreateVssSnapshot SSM document requires that you assign the IAM snapshot role to each instance for which you want to create EBS snapshots. If you don't want to provide additional IAM permissions to your instances for policy or compliance reasons, then you can use the following script.

Before You Begin

Note the following important details about this process:

- This process uses a PowerShell script (CreateVssSnapshotAdvancedScript.ps1) to take snapshots of all volumes on the instances you specify, except root volumes. If you need to take snapshots of root volumes, then you must use the AWSEC2-CreateVssSnapshot SSM document.
- The script calls the AWSEC2-ManageVssIO document twice. The first time with the Action parameter set to Freeze, which pauses all I/O on the instances. The second time, the Action parameter is set to Thaw, which forces I/O to resume.
- Don't attempt to use the AWSEC2-ManageVssIO document without using the CreateVssSnapshotAdvancedScript.ps1 script. A limitation in VSS requires that the Freeze and Thaw actions be called no more than ten seconds apart, and manually calling these actions without the script could result in errors.

To create VSS-enabled EBS snapshots by using the AWSEC2-ManageVssIO SSM document

1. Open AWS Tools for Windows PowerShell and run the following command to specify your credentials. You must either have administrator privileges in Amazon EC2 or have been granted the appropriate permission in IAM. For more information, see Setting Up AWS Systems Manager in the AWS Systems Manager User Guide.

   ```powershell
   Set-AWSCredentials -AccessKey key_name -SecretKey key_name
   ```

2. Execute the following command to set the Region for your PowerShell session. The example uses the us-east-2 Region.

   ```powershell
   Set-DefaultAWSRegion -Region us-east-2
   ```

3. Download the CreateVssSnapshotAdvancedScript.zip file and extract the file contents.

4. Open CreateVssSnapshotAdvancedScript.zip in a text editor, edit the sample call at the bottom of the script with a valid EC2 instance ID, snapshot description, and desired tag values, and then run the script from PowerShell.

If successful, the command populates the list of EBS snapshots with the new snapshots. You can locate these snapshots in the list of EBS snapshots by searching for the tags you specified, or by searching for AppConsistent. If the command execution failed, view the command output for details about why the execution failed. If the command was successfully completed, but a specific volume backup failed, you can troubleshoot the failure in the list of EBS volumes.

Restoring Volumes from VSS-Enabled EBS snapshots

You can use the RestoreVssSnapshotSampleScript.ps1 script to restore volumes on an instance from VSS-enabled EBS snapshots. This script performs the following tasks:
• Stops an instance
• Removes all existing drives from the instance (except the boot volume, if it was excluded)
• Creates new volumes from the snapshots
• Attaches the volumes to the instance by using the device ID tag on the snapshot
• Restarts the instance

**Important**

The following script detaches all volumes attached to an instance, and then creates new volumes from a snapshot. Make sure that you have properly backed-up the instance. The old volumes are not deleted. If you want, you can edit the script to delete the old volumes.

**To restore volumes from VSS-enabled EBS snapshots**

1. Open AWS Tools for Windows PowerShell and run the following command to specify your credentials. You must either have administrator privileges in Amazon EC2 or have been granted the appropriate permission in IAM. For more information, see Setting Up AWS Systems Manager in the AWS Systems Manager User Guide.

   ```powershell
   Set-AWSCredentials -AccessKey key_name -SecretKey key_name
   ```

2. Run the following command to set the Region for your PowerShell session. The example uses the us-east-2 Region.

   ```powershell
   Set-DefaultAWSRegion -Region us-east-2
   ```

3. Download the RestoreVssSnapshotSampleScript.zip file and extract the file contents.

4. Open RestoreVssSnapshotSampleScript.zip in a text editor and edit the sample call at the bottom of the script with a valid EC2 instance ID and EBS snapshot ID, and then run the script from PowerShell.

**Deleting an Amazon EBS Snapshot**

When you delete a snapshot, only the data referenced exclusively by that snapshot is removed. Unique data will not be deleted unless all of the snapshots that reference that data are deleted. Deleting previous snapshots of a volume does not affect your ability to restore volumes from later snapshots of that volume.

Deleting a snapshot of a volume has no effect on the volume. Deleting a volume has no effect on the snapshots made from it.

If you make periodic snapshots of a volume, the snapshots are *incremental*. This means that only the blocks on the device that have changed after your last snapshot are saved in the new snapshot. Even though snapshots are saved incrementally, the snapshot deletion process is designed so that you need to retain only the most recent snapshot in order to restore the volume. Data that was present on a volume, held in an earlier snapshot or series of snapshots, that is subsequently deleted from that volume at a later time, is still considered unique data of the earlier snapshots. This unique data is not deleted from the sequence of snapshots unless all snapshots that reference the unique data are deleted.

Deleting a snapshot might not reduce your organization's data storage costs. Other snapshots might reference that snapshot's data, and referenced data is always preserved. If you delete a snapshot containing data being used by a later snapshot, costs associated with the referenced data are allocated to the later snapshot. For more information about how snapshots store data, see How Incremental Snapshots Work (p. 819) and the example below.

To delete multi-volume snapshots, retrieve all of the snapshots for your multi-volume group using the tag you applied to the group when you created the snapshots. Then, delete the snapshots individually. You will not be prevented from deleting individual snapshots in the multi-volume snapshots group.
In the following diagram, Volume 1 is shown at three points in time. A snapshot has captured each of the first two states, and in the third, a snapshot has been deleted.

- In State 1, the volume has 10 GiB of data. Because Snap A is the first snapshot taken of the volume, the entire 10 GiB of data must be copied.

- In State 2, the volume still contains 10 GiB of data, but 4 GiB have changed. Snap B needs to copy and store only the 4 GiB that changed after Snap A was taken. The other 6 GiB of unchanged data, which are already copied and stored in Snap A, are referenced by Snap B rather than (again) copied. This is indicated by the dashed arrow.

- In state 3, the volume has not changed since State 2, but Snapshot A has been deleted. The 6 GiB of data stored in Snapshot A that were referenced by Snapshot B have now been moved to Snapshot B, as shown by the heavy arrow. As a result, you are still charged for storing 10 GiB of data; 6 GiB of unchanged data preserved from Snap A and 4 GiB of changed data from Snap B.

Example 1: Deleting a Snapshot with Some of its Data Referenced by Another Snapshot
Note that you can't delete a snapshot of the root device of an EBS volume used by a registered AMI. You must first deregister the AMI before you can delete the snapshot. For more information, see Deregistering Your Windows AMI (p. 77).

To delete a snapshot using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Snapshots in the navigation pane.
3. Select a snapshot and then choose Delete from the Actions list.
4. Choose Yes, Delete.

To delete a snapshot using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `delete-snapshot` (AWS CLI)
- `Remove-EC2Snapshot` (AWS Tools for Windows PowerShell)

**Note**
Although you can delete a snapshot that is still in progress, the snapshot must complete before the deletion takes effect. This may take a long time. If you are also at your concurrent snapshot limit (five snapshots in progress), and you attempt to take an additional snapshot, you may get the `ConcurrentSnapshotLimitExceeded` error.

Copying an Amazon EBS Snapshot

With Amazon EBS, you can create point-in-time snapshots of volumes, which we store for you in Amazon S3. After you create a snapshot and it has finished copying to Amazon S3 (when the snapshot status is completed), you can copy it from one AWS Region to another, or within the same Region. Amazon S3 server-side encryption (256-bit AES) protects a snapshot's data in transit during a copy operation. The snapshot copy receives an ID that is different from the ID of the original snapshot.

To copy multi-volume snapshots to another AWS Region, retrieve the snapshots using the tag you applied to the multi-volume snapshots group when you created it. Then individually copy the snapshots to another Region.

For information about copying an Amazon RDS snapshot, see Copying a DB Snapshot in the Amazon RDS User Guide.

If you would like another account to be able to copy your snapshot, you must either modify the snapshot permissions to allow access to that account or make the snapshot public so that all AWS accounts can copy it. For more information, see Sharing an Amazon EBS Snapshot (p. 842).

For pricing information about copying snapshots across AWS Regions and accounts, see Amazon EBS Pricing. Note that snapshot copy operations within a single account and Region do not copy any actual data and therefore are cost-free as long as the encryption status of the snapshot copy does not change.

**Note**
If you copy a snapshot to a new Region, a complete (non-incremental) copy is always created, resulting in additional delay and storage costs.

**Note**
If you copy a snapshot and encrypt it to a new CMK, a complete (non-incremental) copy is always created, resulting in additional delay and storage costs.
Use Cases

- Geographic expansion: Launch your applications in a new AWS Region.
- Migration: Move an application to a new Region, to enable better availability and to minimize cost.
- Disaster recovery: Back up your data and logs across different geographical locations at regular intervals. In case of disaster, you can restore your applications using point-in-time backups stored in the secondary Region. This minimizes data loss and recovery time.
- Encryption: Encrypt a previously unencrypted snapshot, change the key with which the snapshot is encrypted, or, for encrypted snapshots that have been shared with you, create a copy that you own in order to restore a volume from it.
- Data retention and auditing requirements: Copy your encrypted EBS snapshots from one AWS account to another to preserve data logs or other files for auditing or data retention. Using a different account helps prevent accidental snapshot deletions, and protects you if your main AWS account is compromised.

Prerequisites

- You can copy any accessible snapshots that have a completed status, including shared snapshots and snapshots that you have created.
- You can copy AWS Marketplace, VM Import/Export, and AWS Storage Gateway snapshots, but you must verify that the snapshot is supported in the destination Region.

Limits

- Each account can have up to five concurrent snapshot copy requests to a single destination Region.
- User-defined tags are not copied from the source snapshot to the new snapshot. After the copy operation is complete, you can apply user-defined tags to the new snapshot. For more information, see Tagging Your Amazon EC2 Resources (p. 964).
- Snapshots created by the CopySnapshot action have an arbitrary volume ID that should not be used for any purpose.

Incremental Copying Across Regions

The first snapshot copy to another Region is always a full copy. For unencrypted snapshots, each subsequent snapshot copy of the same volume is incremental, meaning that AWS copies only the data that changed since your last snapshot copy to the same destination Region. This allows for faster copying and lower storage costs.

In the case of encrypted snapshots, you must encrypt to the same CMK that was used for previous copies to get incremental copies. The following examples illustrate how this works:

- If you copy an unencrypted snapshot from the US East (N. Virginia) Region to the US West (Oregon) Region, the first snapshot copy is a full copy and subsequent snapshot copies of the same volume transferred between the same Regions are incremental.
- If you copy an encrypted snapshot from the US East (N. Virginia) Region to the US West (Oregon) Region, the first snapshot copy of the volume is a full copy.
- If you encrypt to the same CMK in a subsequent snapshot copy for the same volume between the same Regions, the copy is incremental.
- If you encrypt to a different CMK in a subsequent snapshot copy for the same volume between the same Regions, the copy is a new full copy of the snapshot.

For more information, see Re-Encrypt an Encrypted Snapshot (p. 872).
### Encryption and Snapshot Copying

When you copy a snapshot, you can encrypt the copy or you can specify a CMK different from the original one, and the resulting copied snapshot uses the new CMK. However, changing the encryption status of a snapshot during a copy operation results in a full (not incremental) copy, which might incur greater data transfer and storage charges.

To copy an encrypted snapshot shared from another AWS account, you must have permissions to use the snapshot and the customer master key (CMK) that was used to encrypt the snapshot. When using an encrypted snapshot that was shared with you, we recommend that you re-encrypt the snapshot by copying it using a CMK that you own. This protects you if the original CMK is compromised, or if the owner revokes it, which could cause you to lose access to any encrypted volumes that you created using the snapshot. For more information, see [Sharing an Amazon EBS Snapshot](p. 842).

You apply encryption to EBS snapshot copies by setting the `Encrypted` parameter to `true`. (The `Encrypted` parameter is optional if encryption by default (p. 868) is enabled).

Optionally, you can use `KmsKeyId` to specify a custom key to use to encrypt the snapshot copy. (The `Encrypted` parameter must also be set to `true`, even if encryption by default is enabled.) If `KmsKeyId` is not specified, the key that is used for encryption depends on the encryption state of the source snapshot and its ownership.

The following table describes the encryption outcome for each possible combination of settings.

#### Encryption outcomes: Copying a snapshot

<table>
<thead>
<tr>
<th>Is <code>Encrypted</code> parameter set?</th>
<th>Is encryption by default set?</th>
<th>Source snapshot</th>
<th>Default (no KmsKeyId specified)</th>
<th>Custom (KmsKeyId specified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>Unencrypted snapshot that you own</td>
<td>Unencrypted</td>
<td>N/A</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Encrypted snapshot that you own</td>
<td>Encrypted by same key</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Unencrypted snapshot that is shared with you</td>
<td>Unencrypted</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Encrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK*</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Unencrypted snapshot that you own</td>
<td>Encrypted by default CMK</td>
<td>Encrypted by a specified CMK**</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Encrypted snapshot that you own</td>
<td>Encrypted by same key</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Unencrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Encrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>Is Encrypted parameter set?</td>
<td>Is encryption by default set?</td>
<td>Source snapshot</td>
<td>Default (no KmsKeyId specified)</td>
<td>Custom (KmsKeyId specified)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>-------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Unencrypted snapshot that you own</td>
<td>Encrypted by default CMK</td>
<td>N/A</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Encrypted snapshot that you own</td>
<td>Encrypted by same key</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Unencrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Encrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Unencrypted snapshot that you own</td>
<td>Encrypted by default CMK</td>
<td>Encrypted by a specified CMK</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Encrypted snapshot that you own</td>
<td>Encrypted by same key</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Unencrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Encrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
</tbody>
</table>

* This is the default CMK used for EBS encryption for the AWS account and Region. By default this is a unique AWS managed CMK for EBS, or you can specify a customer managed CMK. For more information, see Default Key for EBS Encryption (p. 867).

** This is a customer managed CMK specified for the copy action. This CMK is used instead of the default CMK for the AWS account and Region.

### Copy a Snapshot

Use the following procedure to copy a snapshot using the Amazon EC2 console.

#### To copy a snapshot using the console

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Snapshots**.
3. Select the snapshot to copy, and then choose **Copy** from the **Actions** list.
4. In the **Copy Snapshot** dialog box, update the following as necessary:

   - **Destination region**: Select the Region where you want to write the copy of the snapshot.
   - **Description**: By default, the description includes information about the source snapshot so that you can identify a copy from the original. You can change this description as necessary.
   - **Encryption**: If the source snapshot is not encrypted, you can choose to encrypt the copy. If you have enabled encryption by default (p. 868), the **Encryption** option is set and cannot be unset.
from the snapshot console. If the Encryption option is set, you can choose to encrypt it to a customer managed CMK by selecting one in the field, described below.

You cannot strip encryption from an encrypted snapshot.

**Note**
If you copy a snapshot and encrypt it to a new CMK, a complete (non-incremental) copy is always created, resulting in additional delay and storage costs.

- **Master Key**: The customer master key (CMK) to be used to encrypt this snapshot. The default key for your account is displayed initially, but you can optionally select from the master keys in your account or type/paste the ARN of a key from a different account. You can create new master encryption keys in the IAM console https://console.aws.amazon.com/iam/.

5. Choose **Copy**.
6. In the **Copy Snapshot** confirmation dialog box, choose **Snapshots** to go to the **Snapshots** page in the Region specified, or choose **Close**.

To view the progress of the copy process, switch to the destination Region, and then refresh the **Snapshots** page. Copies in progress are listed at the top of the page.

To check for failure

If you attempt to copy an encrypted snapshot without having permissions to use the encryption key, the operation fails silently. The error state is not displayed in the console until you refresh the page. You can also check the state of the snapshot from the command line, as in the following example.

```
aws ec2 describe-snapshots --snapshot-id snap-0123abcd
```

If the copy failed because of insufficient key permissions, you see the following message: "StateMessage": "Given key ID is not accessible".

When copying an encrypted snapshot, you must have **DescribeKey** permissions on the default CMK. Explicitly denying these permissions results in copy failure. For information about managing CMK keys, see **Controlling Access to Customer Master Keys**.

**To copy a snapshot using the command line**

You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).

- **copy-snapshot** (AWS CLI)
- **Copy-EC2Snapshot** (AWS Tools for Windows PowerShell)

**Viewing Amazon EBS Snapshot Information**

You can view detailed information about your snapshots.

**To view snapshot information using the console**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose **Snapshots** in the navigation pane.
3. To reduce the list, choose an option from the **Filter** list. For example, to view only your snapshots, choose **Owned By Me**. You can filter your snapshots further by using the advanced search options. Choose the search bar to view the filters available.
4. To view more information about a snapshot, select it.
To view snapshot information using the command line

You can use one of the following commands. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `describe-snapshots` (AWS CLI)
- `Get-EC2Snapshot` (AWS Tools for Windows PowerShell)

Sharing an Amazon EBS Snapshot

By modifying the permissions of a snapshot, you can share it with the AWS accounts that you specify. Users that you have authorized can use the snapshots you share as the basis for creating their own EBS volumes, while your original snapshot remains unaffected. If you choose, you can make your unencrypted snapshots available publicly to all AWS users. You can't make your encrypted snapshots available publicly.

When you share an encrypted snapshot, you must also share the customer managed CMK used to encrypt the snapshot. You can apply cross-account permissions to a customer managed CMK either when it is created or at a later time.

**Important**

When you share a snapshot, you are giving others access to all of the data on the snapshot. Share snapshots only with people with whom you want to share **all** of your snapshot data.

Considerations

The following considerations apply to sharing snapshots:

- Snapshots are constrained to the Region in which they were created. To share a snapshot with another Region, copy the snapshot to that Region. For more information, see Copying an Amazon EBS Snapshot (p. 837).
- If your snapshot uses the longer resource ID format, you can only share it with another account that also supports longer IDs. For more information, see Resource IDs (p. 956).
- AWS prevents you from sharing snapshots that were encrypted with your default CMK. Snapshots that you intend to share must instead be encrypted with a customer managed CMK. For more information, see Creating Keys in the AWS Key Management Service Developer Guide.
- Users of your shared CMK who are accessing encrypted snapshots must be granted permissions to perform the following actions on the key: `kms:DescribeKey`, `kms:CreateGrant`, `GenerateDataKey`, and `kms:ReEncrypt`. For more information, see Controlling Access to Customer Master Keys in the AWS Key Management Service Developer Guide.

Sharing an Unencrypted Snapshot Using the Console

To share a snapshot using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Snapshots in the navigation pane.
3. Select the snapshot and then choose Actions, Modify Permissions.
4. Make the snapshot public or share it with specific AWS accounts as follows:
   - To make the snapshot public, choose Public.

   This option is not valid for encrypted snapshots or snapshots with an AWS Marketplace product code.
• To share the snapshot with one or more AWS accounts, choose Private, enter the AWS account ID (without hyphens) in **AWS Account Number**, and choose Add Permission. Repeat for any additional AWS accounts.

5. Choose **Save**.

**To use an unencrypted snapshot that was privately shared with you**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose **Snapshots** in the navigation pane.
3. Choose the **Private Snapshots** filter.
4. Locate the snapshot by ID or description. You can use this snapshot as you would any other; for example, you can create a volume from the snapshot or copy the snapshot to a different Region.

**Sharing an Encrypted Snapshot Using the Console**

**To share an encrypted snapshot using the console**

2. To change the AWS Region, use the Region selector in the upper-right corner of the page.
3. Choose **Customer managed keys** in the navigation pane.
4. Choose the alias of the customer managed key that you used to encrypt the snapshot.
5. Choose **Add other AWS accounts** and enter the AWS account ID as prompted. To add another account, choose **Add another AWS account** and enter the AWS account ID. When you have added all AWS accounts, choose **Save changes**.
6. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
7. Choose **Snapshots** in the navigation pane.
8. Select the snapshot and then choose **Actions, Modify Permissions**.
9. For each AWS account, enter the AWS account ID in **AWS Account Number** and choose **Add Permission**. When you have added all AWS accounts, choose **Save**.

**To use an encrypted snapshot that was shared with you**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose **Snapshots** in the navigation pane.
3. Choose the **Private Snapshots** filter. Optionally add the **Encrypted** filter.
4. Locate the snapshot by ID or description.
5. Select the snapshot and choose **Actions, Copy**.
6. (Optional) Select a destination Region.
7. The copy of the snapshot is encrypted by the key displayed in **Master Key**. By default, the selected key is your account's default CMK. To select a customer managed CMK, click inside the input box to see a list of available keys.
8. Choose **Copy**.

**Sharing a Snapshot Using the Command Line**

The permissions for a snapshot are specified using the `createVolumePermission` attribute of the snapshot. To make a snapshot public, set the group to all. To share a snapshot with a specific AWS account, set the user to the ID of the AWS account.
To modify snapshot permissions using the command line

Use one of the following commands:

- `modify-snapshot-attribute` (AWS CLI)
- `Edit-EC2SnapshotAttribute` (AWS Tools for Windows PowerShell)

To view snapshot permissions using the command line

Use one of the following commands:

- `describe-snapshot-attribute` (AWS CLI)
- `Get-EC2SnapshotAttribute` (AWS Tools for Windows PowerShell)

For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

Automating the Amazon EBS Snapshot Lifecycle

You can use Amazon Data Lifecycle Manager (Amazon DLM) to automate the creation, retention, and deletion of snapshots taken to back up your Amazon EBS volumes. Automating snapshot management helps you to:

- Protect valuable data by enforcing a regular backup schedule.
- Retain backups as required by auditors or internal compliance.
- Reduce storage costs by deleting outdated backups.

Combined with the monitoring features of Amazon CloudWatch Events and AWS CloudTrail, Amazon DLM provides a complete backup solution for EBS volumes at no additional cost.

Contents

- Understanding Amazon DLM (p. 844)
- Permissions for Amazon DLM (p. 846)
- Permissions for IAM Users (p. 847)
- Limits (p. 847)
- Working with Amazon DLM Using the Console (p. 847)
- Working with Amazon DLM Using the Command Line (p. 849)
- Working with Amazon DLM Using the API (p. 852)
- Working with Amazon DLM to Create and Maintain Multi-Volume Snapshots Using the Console (p. 852)
- Working with Amazon DLM to Create and Maintain Multi-Volume Snapshots Using the CLI (p. 853)
- Monitoring the Snapshot Lifecycle (p. 854)

Understanding Amazon DLM

The following are the key elements that you should understand before you get started with Amazon DLM.

Snapshots

Snapshots are the primary means to back up data from your EBS volumes. To save storage costs, successive snapshots are incremental, containing only the volume data that changed since the previous
snapshot. When you delete one snapshot in a series of snapshots for a volume, only the data unique to that snapshot is removed. The rest of the captured history of the volume is preserved.

For more information, see Amazon EBS Snapshots (p. 818).

**Target Resource Tags**

Amazon DLM uses resource tags to identify the EBS volumes to back up. Tags are customizable metadata that you can assign to your AWS resources (including EBS volumes and snapshots). An Amazon DLM policy (described below) targets a volume for backup using a single tag. Multiple tags can be assigned to a volume if you want to run multiple policies on it.

You can’t use a ‘\’ or ‘=’ character in a tag key.

For more information, see Tagging Your Amazon EC2 Resources (p. 964).

**Snapshot Tags**

Amazon DLM applies the following tags to all snapshots created by a policy, to distinguish them from snapshots created by any other means:

- `aws:dlm:lifecycle-policy-id`
- `aws:dlm:lifecycle-schedule-name`

You can also specify custom tags to be applied to snapshots on creation.

You can’t use a ‘\’ or ‘=’ character in a tag key.

The target tags that Amazon DLM uses to associate volumes with a policy can optionally be applied to snapshots created by the policy.

**Lifecycle Policies**

A lifecycle policy consists of these core settings:

- **Policy type**—Defines valid target resource types and actions a policy can manage. Defaults to `EBS_SNAPSHOT_MANAGEMENT` if not present.
- **Resource type**—The AWS resource managed by the policy. Supported values are EBS volumes and EC2 instances.
- **Target tag**—The tag that must be associated with an EBS volume or an EC2 instance for it to be managed by the policy.
- **Schedule**—Defines how often to create snapshots and the maximum number of snapshots to keep. Snapshot creation starts within an hour of the specified start time. If creating a new snapshot exceeds the maximum number of snapshots to keep for the volume, the oldest snapshot is deleted.

The following considerations apply to lifecycle policies:

- A policy does not begin creating snapshots until you set its activation status to `enabled`. You can configure a policy to be enabled upon creation.
- Snapshots begin to be created by a policy within one hour following the specified start time.
- If you modify a policy by removing or changing its target tag, the EBS volumes with that tag are no longer affected by the policy.
- If you modify the schedule name for a policy, the snapshots created under the old schedule name are no longer affected by the policy.
- If you delete the resource to which a policy applies, the policy no longer manages the previously created snapshots. You must manually delete the snapshots if they are no longer needed.
• You can create multiple policies to back up an EBS volume or an EC2 instance. For example, if an EBS volume has two tags, where tag A is the target for policy A to create a snapshot every 12 hours, and tag B is the target for policy B to create a snapshot every 24 hours, Amazon DLM creates snapshots according to the schedules for both policies.

• When you copy a snapshot created by a policy, the retention schedule is not carried over to the copy. This ensures that Amazon DLM does not delete snapshots that should be retained for a longer period of time.

For example, you could create a policy that manages all EBS volumes with the tag account=Finance, creates snapshots every 24 hours at 0900, and retains the five most recent snapshots. Snapshot creation could start as late as 0959.

Permissions for Amazon DLM

Amazon DLM uses an IAM role to get the permissions that are required to manage snapshots on your behalf. Amazon DLM creates the AWSDataLifecycleManagerDefaultRole role the first time that you create a lifecycle policy using the AWS Management Console. You can also create this role using the following create-default-role command.

```
aws dlm create-default-role
```

Alternatively, you can create a custom IAM role with the required permissions and select it when you create a lifecycle policy.

To create a custom IAM role

1. Create a role with the following permissions.

   ```
   {
     "Version": "2012-10-17",
     "Statement": [
       {
         "Effect": "Allow",
         "Action": [
           "ec2:CreateSnapshot",
           "ec2:CreateSnapshots",
           "ec2:DeleteSnapshot",
           "ec2:DescribeVolumes",
           "ec2:DescribeInstances",
           "ec2:DescribeSnapshots"
         ],
         "Resource": "*"
       },
       {
         "Effect": "Allow",
         "Action": [
           "ec2:CreateTags"
         ],
         "Resource": "arn:aws:ec2::*::snapshot/*"
       }
     ]
   }
   ```

   For more information, see Creating a Role in the IAM User Guide.

2. Add a trust relationship to the role.

   a. In the IAM console, choose Roles.
   b. Select the role you created and then choose Trust relationships.
c. Choose **Edit Trust Relationship**, add the following policy, and then choose **Update Trust Policy**.

```json
{
   "Version": "2012-10-17",
   "Statement": [
   {
      "Effect": "Allow",
      "Principal": {
         "Service": "dlm.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
   }
   ]
}
```

**Permissions for IAM Users**

An IAM user must have the following permissions to use Amazon DLM.

```json
{
   "Version": "2012-10-17",
   "Statement": [
   {
      "Effect": "Allow",
      "Action": ["iam:PassRole", "iam:ListRoles"],
      "Resource": "arn:aws:iam::123456789012:role/AWSDataLifecycleManagerDefaultRole"
   },
   {
      "Effect": "Allow",
      "Action": "dlm:*",
      "Resource": "*"
   }
   ]
}
```

For more information, see [Changing Permissions for an IAM User](https://docs.aws.amazon.com/IAM/latest/UserGuide/id_users.html) in the IAM User Guide.

**Limits**

Your AWS account has the following limits related to Amazon DLM:

- You can create up to 100 lifecycle policies per Region.
- You can add up to 50 tags per resource.
- You can create one schedule per lifecycle policy.

**Working with Amazon DLM Using the Console**

The following examples show how to use Amazon DLM to perform typical procedures to manage the backups of your EBS volumes.

**To create a lifecycle policy**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose **Elastic Block Store, Lifecycle Manager, Create snapshot lifecycle policy**.
3. Provide the following information for your policy as needed:
- **Description**—A description of the policy.
- **Target with tags**—The resource tags that identify the volumes or instances to back up.
- **Schedule Name**—A name for the backup schedule.
- **Create snapshots every n Hours**—The number of hours between policy runs. The supported values are 2, 3, 4, 6, 8, 12, and 24.
- **Snapshot creation start time hh:mm UTC**—The time of day when policy runs are scheduled to start. The policy runs start within an hour after the scheduled time.
- **Retention rule**—The maximum number of snapshots to retain for each volume or instance. The supported range is 1 to 1000. After the limit is reached, the oldest snapshot is deleted when a new one is created.
- **Copy tags**—Copy all user-defined tags on a source volume to snapshots of the volume created by this policy.
- **Tag created snapshots**—The resource tags to apply to the snapshots that are created. These tags are in addition to the tags applied by Amazon DLM. You can also choose variable tags that can automatically tag all of your snapshots with the corresponding instance-id or timestamp.
- **IAM role**—An IAM role that has permissions to create, delete, and describe snapshots, and to describe volumes. AWS provides a default role, **AWSDataLifecycleManagerDefaultRole**, or you can create a custom IAM role.
- **Policy status after creation**—Choose **Enable policy** to start the policy runs at the next scheduled time or **Disable policy** to prevent the policy from running.

4. Choose **Create Policy**.

**To display a lifecycle policy**

1. Open the Amazon EC2 console at **https://console.aws.amazon.com/ec2/**.
2. In the navigation pane, choose **Elastic Block Store, Lifecycle Manager**.
3. Select a lifecycle policy from the list. The **Details** tab displays the following information about the policy:
   - **Policy ID**
   - **Resource type**
   - **Policy type**
   - **Date created**
   - **Date modified**
   - **Target resource tags**
   - **Rule summary**
   - **Description**
   - **Policy state**
   - **Tags added to snapshots**

**To modify a lifecycle policy**

1. Open the Amazon EC2 console at **https://console.aws.amazon.com/ec2/**.
2. In the navigation pane, choose **Elastic Block Store, Lifecycle Manager**.
3. Select a lifecycle policy from the list.
4. Choose **Actions, Modify policy**.
5. In an existing lifecycle policy, you can modify the following policy values:
   - **Description**—A description of the policy.
• **Target with tags**—The resource tags that identify the volumes or instances to back up.
• **Schedule Name**—A name for the backup schedule.
• **Create snapshots every n Hours**—The number of hours between policy runs. The supported values are 2, 3, 4, 6, 8, 12, and 24.
• **Snapshot creation start time hh:mm UTC**—The time of day when policy runs are scheduled to start. The policy runs start within an hour after the scheduled time.
• **Retention rule**—The maximum number of snapshots to retain for each volume or instance. The supported range is 1 to 1000. After the limit is reached, the oldest snapshot is deleted when a new one is created.
• **Copy tags**—Copy all user-defined tags on a source volume to snapshots of the volume created by this policy.
• **Tag created snapshots**—The resource tags to apply to the snapshots that are created. These tags are in addition to the tags applied by Amazon DLM. You can also choose variable tags that can automatically tag all of your snapshots with the corresponding instance-id or timestamp.
• **IAM role**—An IAM role that has permissions to create, delete, and describe snapshots, and to describe volumes. AWS provides a default role, `AWSDataLifecycleManagerDefaultRole`, or you can create a custom IAM role.
• **Policy status after creation**—Choose `Enable policy` to start the policy runs at the next scheduled time or `Disable policy` to prevent the policy from running.

**To delete a lifecycle policy**
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Elastic Block Store, Lifecycle Manager**.
3. Select a lifecycle policy from the list.
4. Choose **Actions, Delete policy**.

**Working with Amazon DLM Using the Command Line**

The following examples show how to use Amazon DLM to perform typical procedures to manage the backups of your EBS volumes.

**Example Example: Create a lifecycle policy**

Use the `create-lifecycle-policy` command to create a lifecycle policy. To simplify the syntax, this example references a JSON file, `policyDetails.json`, that includes the policy details.

```
aws dlm create-lifecycle-policy --description "My first policy" --state ENABLED --
execution-role-arn arn:aws:iam::12345678910:role/AWSDataLifecycleManagerDefaultRole --
policy-details file://policyDetails.json
```

The following is an example of the `policyDetails.json` file.

```json
{
    "ResourceTypes": [
        "VOLUME"
    ],
    "TargetTags": [
        {
            "Key": "costcenter",
            "Value": "115"
        }
    ],
    "Schedules":
```
Upon success, the command returns the ID of the newly created policy. The following is example output.

```
{
   "PolicyId": "policy-0123456789abcdef0"
}
```

**Example Example: Display a lifecycle policy**

Use the `get-lifecycle-policy` command to display information about a lifecycle policy.

```
aws dlm get-lifecycle-policy --policy-id policy-0123456789abcdef0
```

The following is example output. It includes the information that you specified, plus metadata inserted by AWS.

```
{
   "Policy":{
      "Description": "My first policy",
      "DateCreated": "2018-05-15T00:16:21+0000",
      "State": "ENABLED",
      "ExecutionRoleArn": "arn:aws:iam::210774411744:role/AWSDataLifecycleManagerDefaultRole",
      "PolicyId": "policy-0123456789abcdef0",
      "DateModified": "2018-05-15T00:16:22+0000",
      "PolicyDetails": {
         "PolicyType":"EBS_SNAPSHOT_MANAGEMENT",
         "ResourceTypes": ["VOLUME"],
         "TargetTags": ["costcenter"],
         "Schedules": []
      }
   }
}
```
Example To modify a lifecycle policy

Use the `update-lifecycle-policy` command to modify the information in a lifecycle policy. To simplify the syntax, this example references a JSON file, `policyDetailsUpdated.json`, that includes the policy details.

```bash
aws dlm update-lifecycle-policy --state DISABLED --execution-role-arn arn:aws:iam::12345678910:role/AWSDataLifecycleManagerDefaultRole --policy-details file://policyDetailsUpdated.json
```

The following is an example of the `policyDetailsUpdated.json` file.

```json
{
   "ResourceTypes": [
      "VOLUME"
   ],
   "TargetTags": [
      {
         "Key": "costcenter",
         "Value": "120"
      }
   ],
   "Schedules": [
      {
         "Name": "DailySnapshots",
         "TagsToAdd": [
            {
               "Key": "type",
               "Value": "myDailySnapshot"
            }
         ],
         "CreateRule": {
            "Interval": 12,
            "IntervalUnit": "HOURS",
            "Times": [
               "15:00"
            ]
         },
         "RetainRule": {
            "Count": 5
         }
      }
   ]
}
```
To view the updated policy, use the `get-lifecycle-policy` command. You can see that the state, the value of the tag, the snapshot interval, and the snapshot start time were changed.

**Example Example: Delete a lifecycle policy**

Use the `delete-lifecycle-policy` command to delete a lifecycle policy and free up the target tags specified in the policy for reuse.

```
aws dlm delete-lifecycle-policy --policy-id policy-0123456789abcdef0
```

**Working with Amazon DLM Using the API**

The Amazon Data Lifecycle Manager API Reference provides descriptions and syntax for each of the actions and data types for the Amazon DLM Query API.

Alternatively, you can use one of the AWS SDKs to access an API that's tailored to the programming language or platform that you're using. For more information, see AWS SDKs.

**Working with Amazon DLM to Create and Maintain Multi-Volume Snapshots Using the Console**

You can create lifecycle policies to automate the creation and deletion of multi-volume snapshots by using the AWS Management Console, from the AWS CLI, or from the Data Lifecycle Management (DLM) APIs.

**Automate multi-volume snapshots using the console**

1. Sign in to the AWS Management Console and open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Elastic Block Store. Then choose Lifecycle Manager and Create snapshot lifecycle policy.
3. Provide the following information for your policy, as needed:
   - **Description**—A description of the policy.
   - **Target with tags**—The resource tags that identify the volumes or instances to back up.
   - **Schedule Name**—A name for the backup schedule.
   - **Create snapshots every n Hours**—The number of hours between policy runs. The supported values are 2, 3, 4, 6, 8, 12, and 24.
   - **Snapshot creation start time hh:mm UTC**—The time of day when policy runs are scheduled to start. The policy runs start within an hour after the scheduled time.
   - **Retention rule**—The maximum number of snapshots to retain for each volume or instance. The supported range is 1 to 1000. After the limit is reached, the oldest snapshot is deleted when a new one is created.
   - **Copy tags**—Copy all user-defined tags on a source volume to snapshots of the volume created by this policy.
   - **Tag created snapshots**—The resource tags to apply to the snapshots that are created. These tags are in addition to the tags applied by Amazon DLM. You can also choose variable tags that can automatically tag all of your snapshots with the corresponding instance-id or timestamp.
- **IAM role**—An IAM role that has permissions to create, delete, and describe snapshots, and to describe volumes. AWS provides a default role, AWSDataLifecycleManagerDefaultRole, or you can create a custom IAM role.

- **Policy status after creation**—Choose Enable policy to start the policy runs at the next scheduled time or Disable policy to prevent the policy from running.

4. Choose Create Policy.

### Working with Amazon DLM to Create and Maintain Multi-Volume Snapshots Using the CLI

The following examples show how to use Amazon DLM to automate creation and deletion of multi-volume snapshots using the AWS CLI.

#### Example Example: Create a lifecycle policy

Use the create-lifecycle-policy command to create a lifecycle policy. To simplify the syntax, this example references a JSON file, policyDetails.json, which includes the policy details.

```
```

The following is an example of the multi-volume-policy.json file.

```json
{
  "ResourceTypes": [
    "INSTANCE"
  ],
  "TargetTags": [
    {
      "Key": "costcenter",
      "Value": "115"
    }
  ],
  "Schedules": [
    {
      "Name": "DailySnapshots",
      "TagsToAdd": [
        {
          "Key": "type",
          "Value": "Daily-Multi-Volume Snapshots"
        }
      ],
      "VariableTags": [
        {
          "Key": "timestamp",
          "Value": "$(timestamp)"
        },
        {
          "Key": "instance-id",
          "Value": "$(instance-id)"
        }
      ],
      "Interval": 24,
      "IntervalUnit": "HOURS",
      "Times": [
        "03:00"
      ]
    }
  ]
}
```
Upon success, the command returns the ID of the newly created policy. The following is example output.

```
{
    "PolicyId": "policy-0123456789abcdef0"
}
```

### Monitoring the Snapshot Lifecycle

You can use the following features to monitor the lifecycle of your snapshots.

#### Console and AWS CLI

You can view your lifecycle policies using the Amazon EC2 console or the AWS CLI. Each snapshot created by a policy has a timestamp and policy-related tags. You can filter snapshots using tags to verify that your backups are being created as you intend. For information about viewing lifecycle policies using the console, see To display a lifecycle policy (p. 848). For information about displaying information about lifecycle policies using the CLI, see Example: Display a lifecycle policy (p. 850).

#### CloudWatch Events

Amazon EBS and Amazon DLM emit events related to lifecycle policy actions. You can use AWS Lambda and Amazon CloudWatch Events to handle event notifications programmatically. For more information, see the Amazon CloudWatch Events User Guide.

The following events are available:

- **createSnapshot**—An Amazon EBS event emitted when a CreateSnapshot action succeeds or fails. For more information, see Amazon CloudWatch Events for Amazon EBS (p. 907).
- **DLM Policy State Change**—A Amazon DLM event emitted when a lifecycle policy enters an error state. The event contains a description of what caused the error. The following is an example of an event when the permissions granted by the IAM role are insufficient:

```
{
    "version": "0",
    "id": "01234567-0123-0123-0123-0123456789ab",
    "detail-type": "DLM Policy State Change",
    "source": "aws.dlm",
    "account": "123456789012",
    "region": "us-east-1",
    "resources": [
        "arn:aws:dlm:us-east-1:123456789012:policy/policy-0123456789abcdef"
    ],
    "detail": {
        "state": "ERROR",
        "cause": "Role provided does not have sufficient permissions",
        "policy_id": "arn:aws:dlm:us-east-1:123456789012:policy/policy-0123456789abcdef"
    }
}
```
The following is an example of an event when a limit is exceeded:

```
{
    "version": "0",
    "id": "01234567-0123-0123-0123-0123456789ab",
    "detail-type": "DLM Policy State Change",
    "source": "aws.dlm",
    "account": "123456789012",
    "time": "2018-05-25T13:12:22Z",
    "region": "us-east-1",
    "resources": [
        "arn:aws:dlm:us-east-1:123456789012:policy/policy-0123456789abcdef"
    ],
    "detail": {
        "state": "ERROR",
        "cause": "Maximum allowed active snapshot limit exceeded",
        "policy_id": "arn:aws:dlm:us-east-1:123456789012:policy/policy-0123456789abcdef"
    }
}
```

**AWS CloudTrail**

With AWS CloudTrail, you can track user activity and API usage to demonstrate compliance with internal policies and regulatory standards. For more information, see the [AWS CloudTrail User Guide](https://docs.aws.amazon.com/AmazonCloudTrail/latest/userguide/).

**AWS CloudFormation**

When deploying resource stacks with AWS CloudFormation, you can include Amazon DLM policies in your AWS CloudFormation templates. For more information, see [Amazon Data Lifecycle Manager Resource Types Reference](https://docs.aws.amazon.com/dlm/latest/userguide/Amazon-Data-Lifecycle-Manager-Resource-Types-Reference.html).

**Amazon EBS Data Services**

Amazon EBS provides the following data services.

**Data Services**

- Amazon EBS Elastic Volumes (p. 855)
- Amazon EBS Encryption (p. 865)

**Amazon EBS Elastic Volumes**

With Amazon EBS Elastic Volumes, you can increase the volume size, change the volume type, or adjust the performance of your EBS volumes. If your instance supports Elastic Volumes, you can do so without detaching the volume or restarting the instance. This enables you to continue using your application while the changes take effect.

There is no charge to modify the configuration of a volume. You are charged for the new volume configuration after volume modification starts. For more information, see the [Amazon EBS Pricing page](https://aws.amazon.com/pricing/).

**Contents**

- Requirements When Modifying Volumes (p. 856)
- Requesting Modifications to Your EBS Volumes (p. 857)
- Monitoring the Progress of Volume Modifications (p. 860)
- Extending a Windows File System After Resizing a Volume (p. 863)
Requirements When Modifying Volumes

The following requirements and limitations apply when you modify an Amazon EBS volume. To learn more about the general requirements for EBS volumes, see Constraints on the Size and Configuration of an EBS Volume (p. 798).

Amazon EC2 Instance Support

Elastic Volumes are supported on the following instances:

- All current-generation instances (p. 120)
- Previous-generation instance families C1, C3, CC2, CR1, G2, I2, M1, M3, and R3

If your instance type does not support Elastic Volumes, see Modifying an EBS Volume If Elastic Volumes Is Unsupported (p. 860).

Requirements for Windows Volumes

By default, Windows initializes volumes with a master boot record (MBR) partition table. Because MBR supports only volumes smaller than 2 TiB (2,048 GiB), Windows prevents you from resizing MBR volumes beyond this limit. In such a case, the Extend Volume option is disabled in the Windows Disk Management utility. If you use the AWS Management Console or AWS CLI to create an MBR-partitioned volume that exceeds the size limit, Windows cannot detect or use the additional space. For requirements affecting Linux volumes, see Requirements for Linux Volumes in the Amazon EC2 User Guide for Linux Instances.

To overcome this limitation, you can create a new, larger volume with a GUID partition table (GPT) and copy over the data from the original MBR volume.

To create a GPT volume

1. Create a new, empty volume of the desired size in the Availability Zone of the EC2 instance and attach it to your instance.
   
   **Note**
   The new volume must not be a volume restored from a snapshot.

2. Log in to your Windows system and open Disk Management (diskmgmt.exe).
3. Open the context (right-click) menu for the new disk and choose Online.
4. In the Initialize Disk window, select the new disk and choose GPT (GUID Partition Table), OK.
5. When initialization is complete, copy the data from the original volume to the new volume, using a tool such as robocopy or teracopy.
6. In Disk Management, change the drive letters to appropriate values and take the old volume offline.
7. In the Amazon EC2 console, detach the old volume from the instance, reboot the instance to verify that it functions properly, and delete the old volume.

Limitations

- The new volume size cannot exceed the supported volume capacity. For more information, see Constraints on the Size and Configuration of an EBS Volume (p. 798).
- If the volume was attached before November 3, 2016 23:40 UTC, you must initialize Elastic Volumes support. For more information, see Initializing Elastic Volumes Support (p. 858).
- If you are using an unsupported previous-generation instance type, or if you encounter an error while attempting a volume modification, see Modifying an EBS Volume If Elastic Volumes Is Unsupported (p. 860).
• A gp2 volume that is attached to an instance as a root volume cannot be modified to an st1 or sc1 volume. If detached and modified to st1 or sc1, it cannot be attached to an instance as the root volume.

• A gp2 volume cannot be modified to an st1 or sc1 volume if the requested volume size is below the minimum size for st1 and sc1 volumes.

• In some cases, you must detach the volume or stop the instance for modification to proceed. If you encounter an error message while attempting to modify an EBS volume, or if you are modifying an EBS volume attached to a previous-generation instance type, take one of the following steps:
  • For a non-root volume, detach the volume from the instance, apply the modifications, and then re-attach the volume.
  • For a root (boot) volume, stop the instance, apply the modifications, and then restart the instance.

• After provisioning over 32,000 IOPS on an existing io1 volume, you may need to do one of the following to see the full performance improvements:
  • Detach and attach the volume.
  • Restart the instance.

• Decreasing the size of an EBS volume is not supported. However, you can create a smaller volume and then migrate your data to it using an application-level tool such as robocopy.

• Modification time is increased if you modify a volume that has not been fully initialized. For more information see Initializing Amazon EBS Volumes (p. 891).

• After modifying a volume, wait at least six hours and ensure that the volume is in the in-use or available state before making additional modifications to the same volume.

• While m3.medium instances fully support volume modification, m3.large, m3.xlarge, and m3.2xlarge instances might not support all volume modification features.

### Requesting Modifications to Your EBS Volumes

With Elastic Volumes, you can dynamically modify the size, performance, and volume type of your Amazon EBS volumes without detaching them.

Use the following process when modifying a volume:

1. (Optional) Before modifying a volume that contains valuable data, it is a best practice to create a snapshot of the volume in case you need to roll back your changes. For more information, see Creating Amazon EBS Snapshots (p. 822).

2. Request the volume modification.

3. Monitor the progress of the volume modification. For more information, see Monitoring the Progress of Volume Modifications (p. 860).

4. If the size of the volume was modified, extend the volume's file system to take advantage of the increased storage capacity. For more information, see Extending a Windows File System After Resizing a Volume (p. 863).

### Contents

- Modifying an EBS Volume Using Elastic Volumes (Console) (p. 857)
- Modifying an EBS Volume Using Elastic Volumes (AWS CLI) (p. 858)
- Initializing Elastic Volumes Support (If Needed) (p. 858)
- Modifying an EBS Volume If Elastic Volumes Is Unsupported (p. 860)

### Modifying an EBS Volume Using Elastic Volumes (Console)

Use the following procedure to modify an EBS volume.
To modify an EBS volume using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Volumes, select the volume to modify, and then choose Actions, Modify Volume.
3. The Modify Volume window displays the volume ID and the volume's current configuration, including type, size, and IOPS. You can change any or all of these settings in a single action. Set new configuration values as follows:
   - To modify the type, choose a value for Volume Type.
   - To modify the size, enter an allowed integer value for Size.
   - If you chose Provisioned IOPS SSD (io1) as the volume type, enter an allowed integer value for IOPS.
4. After you have finished changing the volume settings, choose Modify. When prompted for confirmation, choose Yes.
5. Modifying volume size has no practical effect until you also extend the volume's file system to make use of the new storage capacity. For more information, see Extending a Windows File System After Resizing a Volume (p. 863).
6. If you increase the size of an NVMe volume on an instance that does not have the AWS NVMe drivers, you must reboot the instance to enable Windows to see the new volume size. For more information about installing the AWS NVMe drivers, see AWS NVMe Drivers for Windows Instances (p. 446).

Modifying an EBS Volume Using Elastic Volumes (AWS CLI)

Use the modify-volume command to modify one or more configuration settings for a volume. For example, if you have a volume of type gp2 with a size of 100 GiB, the following command changes its configuration to a volume of type io1 with 10,000 IOPS and a size of 200 GiB.

```
aws ec2 modify-volume --volume-type io1 --iops 10000 --size 200 --volume-id vol-11111111111111111
```

The following is example output:

```
{
    "VolumeModification": {
        "TargetSize": 200,
        "TargetVolumeType": "io1",
        "ModificationState": "modifying",
        "VolumeId": "vol-11111111111111111",
        "TargetIops": 10000,
        "StartTime": "2017-01-19T22:21:02.959Z",
        "Progress": 0,
        "OriginalVolumeType": "gp2",
        "OriginalIops": 300,
        "OriginalSize": 100
    }
}
```

Modifying volume size has no practical effect until you also extend the volume's file system to make use of the new storage capacity. For more information, see Extending a Windows File System After Resizing a Volume (p. 863).

Initializing Elastic Volumes Support (If Needed)

Before you can modify a volume that was attached to an instance before November 3, 2016 23:40 UTC, you must initialize volume modification support using one of the following actions:
• Detach and attach the volume
• Stop and start the instance

Use one of the following procedures to determine whether your instances are ready for volume modification.

**To determine whether your instances are ready using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. On the navigation pane, choose **Instances**.
3. Choose the **Show/Hide Columns** icon (the gear). Select the **Launch Time** and **Block Devices** attributes and then choose **Close**.
4. Sort the list of instances by the **Launch Time** column. For instances that were started before the cutoff date, check when the devices were attached. In the following example, you must initialize volume modification for the first instance because it was started before the cutoff date and its root volume was attached before the cutoff date. The other instances are ready because they were started after the cutoff date.

<table>
<thead>
<tr>
<th>Instance ID</th>
<th>Launch Time</th>
<th>Block Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>i-e905622e</td>
<td>February 25, 2016 at 1:49:35 PM UTC-8</td>
<td>/dev/xvda=vol-e6b46410:attached:2016-11-01</td>
</tr>
<tr>
<td>i-719f99a8</td>
<td>December 8, 2016 at 2:21:51 PM UTC-8</td>
<td>/dev/xvda=vol-bad60e7a:attached:2016-11-01</td>
</tr>
<tr>
<td>i-006b02c1b78381e57</td>
<td>May 17, 2017 at 1:52:52 PM UTC-7</td>
<td>/dev/sda1=vol-0de9250441c730241:attached:2016-11-01</td>
</tr>
<tr>
<td>i-e3d172ed</td>
<td>May 17, 2017 at 2:48:54 PM UTC-7</td>
<td>/dev/sda1=vol-04c34d0b:attached:2016-11-01</td>
</tr>
</tbody>
</table>

**To determine whether your instances are ready using the CLI**

Use the following `describe-instances` command to determine whether the volume was attached before November 3, 2016 23:40 UTC.

```
aws ec2 describe-instances --query "Reservations[*].Instances[*].
[InstanceId,LaunchTime<='2016-11-01',BlockDeviceMappings[*][Ebs.AttachTime<='2016-11-01']]
" --output text
```

The first line of the output for each instance shows its ID and whether it was started before the cutoff date (True or False). The first line is followed by one or more lines that show whether each EBS volume was attached before the cutoff date (True or False). In the following example output, you must initialize volume modification for the first instance because it was started before the cutoff date and its root volume was attached before the cutoff date. The other instances are ready because they were started after the cutoff date.

```
i-e905622e True
i-719f99a8 False
i-006b02c1b78381e57 False
i-e3d172ed False
```

859
Modifying an EBS Volume If Elastic Volumes Is Unsupported

If you are using a supported instance type, you can use Elastic Volumes to dynamically modify the size, performance, and volume type of your Amazon EBS volumes without detaching them.

If you cannot use Elastic Volumes but you need to modify the root (boot) volume, you must stop the instance, modify the volume, and then restart the instance.

After the instance has started, you can check the file system size to see if your instance recognizes the larger volume space.

If the size does not reflect your newly expanded volume, you must extend the file system of your device so that your instance can use the new space. For more information, see Extending a Windows File System After Resizing a Volume (p. 863).

You may have to bring the volume online in order to use it. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804). You do not need to reformat the volume.

Monitoring the Progress of Volume Modifications

When you modify an EBS volume, it goes through a sequence of states. The volume enters the modifying state, the optimizing state, and finally the completed state. At this point, the volume is ready to be further modified.

Note

Rarely, a transient AWS fault can result in a failed state. This is not an indication of volume health; it merely indicates that the modification to the volume failed. If this occurs, retry the volume modification.

While the volume is in the optimizing state, your volume performance is in between the source and target configuration specifications. Transitional volume performance will be no less than the source volume performance. If you are downgrading IOPS, transitional volume performance is no less than the target volume performance.

Volume modification changes take effect as follows:

- Size changes usually take a few seconds to complete and take effect after a volume is in the Optimizing state.
- Performance (IOPS) changes can take from a few minutes to a few hours to complete and are dependent on the configuration change being made.
- It may take up to 24 hours for a new configuration to take effect, and in some cases more, such as when the volume has not been fully initialized. Typically, a fully used 1-TiB volume takes about 6 hours to migrate to a new performance configuration.

Use one of the following methods to monitor the progress of a volume modification.

Contents

- Monitoring the Progress of a Volume Modification (Console) (p. 860)
- Monitoring the Progress of a Volume Modification (AWS CLI) (p. 861)
- Monitoring the Progress of a Volume Modification (CloudWatch Events) (p. 862)

Monitoring the Progress of a Volume Modification (Console)

Use the following procedure to view the progress of one or more volume modifications.

To monitor progress of a modification using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Volumes**.
3. Select the volume. The volume status is displayed in the **State** column and in the **State** field of the details pane. In this example, the modification state is **completed**.
4. Open the information icon next to the **State** field to display before and after information about the most recent modification action, as shown in this example.

### Monitoring the Progress of a Volume Modification (AWS CLI)

Use the `describe-volumes-modifications` command to view the progress of one or more volume modifications. The following example describes the volume modifications for two volumes.

```
aws ec2 describe-volumes-modifications --volume-ids vol-11111111111111111 vol-22222222222222222
```

In the following example output, the volume modifications are still in the **modifying** state.

```json
{
  "VolumesModifications": [
    {
      "TargetSize": 200,
```
The next example describes all volumes with a modification state of either optimizing or completed, and then filters and formats the results to show only modifications that were initiated on or after February 1, 2017:

```
aws ec2 describe-volumes-modifications --filters Name=modification-state,Values="optimizing","completed" --query "VolumesModifications[?StartTime>='2017-02-01'].{ID:VolumeId,STATE:ModificationState}"
```

The following is example output with information about two volumes:

```
[
  {
    "STATE": "optimizing",
    "ID": "vol-06397e7a0eEXAMPLE"
  },
  {
    "STATE": "completed",
    "ID": "vol-ba74e18c2aEXAMPLE"
  }
]
```

### Monitoring the Progress of a Volume Modification (CloudWatch Events)

With CloudWatch Events, you can create a notification rule for volume modification events. You can use your rule to generate a notification message using Amazon SNS or to invoke a Lambda function in response to matching events.

To monitor progress of a modification using CloudWatch Events

2. Choose Events, Create rule.
3. For Build event pattern to match events by service, choose Custom event pattern.
4. For Build custom event pattern, replace the contents with the following and choose Save.

```javascript
{
}
```
Extending a Windows File System After Resizing a Volume

After you increase the size of an EBS volume, use the Windows Disk Management utility to extend the disk size to the new size of the volume. You can begin resizing the file system as soon as the volume enters the optimizing state.

For more information about extending a file system on Linux, see Extending a Linux File System After Resizing a Volume in the Amazon EC2 User Guide for Linux Instances.

To extend a Windows file system

1. Before extending a file system that contains valuable data, it is a best practice to create a snapshot of the volume that contains it in case you need to roll back your changes. For more information, see Creating Amazon EBS Snapshots (p. 822).
2. Log in to your Windows instance using Remote Desktop.
3. In the Run dialog, type diskmgmt.msc and press Enter. The Disk Management utility opens.
4. On the **Disk Management** menu, choose **Action, Rescan Disks**.

5. Open the context (right-click) menu for the expanded drive and choose **Extend Volume**.

6. In the **Extend Volume** wizard, choose **Next**. For **Select the amount of space in MB**, enter the number of megabytes by which to extend the volume. Generally, you specify the maximum available
space. The highlighted text under **Selected** is the amount of space that is added, not the final size the volume will have. Complete the wizard.

7. If you increase the size of an NVMe volume on an instance that does not have the AWS NVMe driver, you must reboot the instance to enable Windows to see the new volume size. For more information about installing the AWS NVMe driver, see **AWS NVMe Drivers for Windows Instances** (p. 446).

### Amazon EBS Encryption

Amazon EBS encryption offers a straightforward encryption solution for your EBS resources that doesn't require you to build, maintain, and secure your own key management infrastructure. It uses AWS Key Management Service (AWS KMS) customer master keys (CMK) when creating encrypted volumes and snapshots.

Encryption operations occur on the servers that host EC2 instances, ensuring the security of both data-at-rest and data-in-transit between an instance and its attached EBS storage.

#### Contents

- How EBS Encryption Works (p. 866)
- Requirements (p. 866)
- Default Key for EBS Encryption (p. 867)
- Encryption by Default (p. 868)
- Encrypting EBS Resources (p. 868)
- Encryption Scenarios (p. 869)
- Setting Encryption Defaults Using the API and CLI (p. 874)
How EBS Encryption Works

You can encrypt both the boot and data volumes of an EC2 instance. When you create an encrypted EBS volume and attach it to a supported instance type, the following types of data are encrypted:

- Data at rest inside the volume
- All data moving between the volume and the instance
- All snapshots created from the volume
- All volumes created from those snapshots

EBS encrypts your volume with a data key using the industry-standard AES-256 algorithm. Your data key is stored on-disk with your encrypted data, but not before EBS encrypts it with your CMK. Your data key never appears on disk in plaintext. The same data key is shared by snapshots of the volume and any subsequent volumes created from those snapshots. For more information, see Data Keys in the AWS Key Management Service Developer Guide.

Amazon EBS works with AWS KMS to encrypt and decrypt your EBS volumes as follows:

1. Amazon EBS sends a CreateGrant request to AWS KMS, so that it can decrypt the data key.
2. Amazon EBS sends a GenerateDataKeyWithoutPlaintext request to AWS KMS, specifying the CMK to use to encrypt the volume.
3. AWS KMS generates a new data key, encrypts it under the specified CMK, and sends the encrypted data key to Amazon EBS to be stored with the volume metadata.
4. When you attach an encrypted volume to an instance, Amazon EBS sends a Decrypt request to AWS KMS, specifying the encrypted data key.
5. AWS KMS decrypts the encrypted data key and sends the decrypted data key to Amazon EBS.
6. Amazon EBS uses the plaintext data key in hypervisor memory to encrypt disk I/O to the volume. The plaintext data key persists in memory as long as the volume is attached to the instance.

For more information, see How Amazon Elastic Block Store (Amazon EBS) Uses AWS KMS and AWS KMS Log File Entries in the AWS Key Management Service Developer Guide.

Requirements

Before you begin, verify that the following requirements are met.

Supported Volume Types

Encryption is supported by all EBS volume types. You can expect the same IOPS performance on encrypted volumes as on unencrypted volumes, with a minimal effect on latency. You can access encrypted volumes the same way that you access unencrypted volumes. Encryption and decryption are handled transparently, and they require no additional action from you or your applications.

Supported Instance Types

Amazon EBS encryption is available on the instance types listed below. You can attach both encrypted and unencrypted volumes to these instance types simultaneously.

- General purpose: M3, M4, M5, M5a, M5ad, M5d, T2, T3, and T3a
- Compute optimized: C3, C4, C5, C5d, and C5n
- Memory optimized: cr1.8xlarge, R3, R4, R5, R5a, R5ad, R5d, u-6tb1.metal, u-9tb1.metal, u-12tb1.metal, X1, X1e, and z1d
• Storage optimized: D2, h1.2xlarge, h1.4xlarge, I2, and I3
• Accelerated computing: F1, G2, G3, G4, P2, and P3

Permissions for IAM Users

When you configure a CMK as the default key for EBS encryption, the default key policy allows any IAM user with access to the required KMS actions to use this key to encrypt or decrypt EBS resources. You must grant IAM users permission to call the following actions in order to use EBS encryption:

- kms:CreateGrant
- kms:Decrypt
- kms:DescribeKey
- kms:GenerateDataKeyWithoutPlainText
- kms:ReEncrypt

To follow the principal of least privilege, do not allow full access to kms:CreateGrant. Instead, allow the user to create grants on the CMK only when the grant is created on the user's behalf by an AWS service, as shown in the following example:

```json
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": "kms:CreateGrant",
         "Resource": [
            "arn:aws:kms:us-east-2:123456789012:key/abcd1234-a123-456d-a12b-a123b4cd56ef"
         ],
         "Condition": {
            "Bool": {
               "kms:GrantIsForAWSResource": true
            }
         }
      }
   ]
}
```

For more information, see Default Key Policy in the AWS Key Management Service Developer Guide.

Default Key for EBS Encryption

Amazon EBS automatically creates a unique AWS managed CMK in each Region where you store AWS resources. This key has the alias alias/aws/ebs. By default, Amazon EBS uses this key for encryption. Alternatively, you can specify a customer managed CMK that you created as the default key for EBS encryption. Using your own CMK gives you more flexibility, including the ability to create, rotate, and disable keys.

To configure the default key for EBS encryption for a Region

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region.
3. Choose Account Attributes, Settings.
4. Choose Change the default key and then choose an available key.
5. Choose Update.

**Encryption by Default**

You can configure your AWS account to enforce the encryption of the new EBS volumes and snapshot copies that you create. For example, Amazon EBS encrypts the EBS volumes created when you launch an instance and the snapshots that you copy from an unencrypted snapshot. For examples of transitioning from unencrypted to encrypted EBS resources, see Encrypting Unencrypted Resources (p. 869).

Encryption by default has no effect on existing EBS volumes or snapshots.

**Considerations**

- Encryption by default is a Region-specific setting. If you enable it for a Region, you cannot disable it for individual volumes or snapshots in that Region.
- When you enable encryption by default, you can launch an instance only if the instance type supports EBS encryption. For more information, see Supported Instance Types (p. 866).
- When migrating servers using AWS Server Migration Service (SMS), do not turn on encryption by default. If encryption by default is already on and you are experiencing delta replication failures, turn off this feature.

**To enable encryption by default for a Region**

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region.
3. From the navigation pane, select EC2 Dashboard.
4. In the upper-right corner of the page, choose Account Attributes, Settings.
5. Under EBS Storage, select Always encrypt new EBS volumes.
6. Choose Update.

You cannot change the CMK that is associated with an existing snapshot or encrypted volume. However, you can associate a different CMK during a snapshot copy operation so that the resulting copied snapshot is encrypted by the new CMK.

**Encrypting EBS Resources**

You encrypt EBS volumes by enabling encryption, either using encryption by default (p. 868) or by enabling encryption when you create a volume that you want to encrypt.

When you encrypt a volume, you can specify the CMK to use to encrypt the volume. If you do not specify a CMK, the key that is used for encryption depends on the encryption state of the source snapshot and its ownership. For more information, see the encryption outcomes table (p. 873).

You cannot change the CMK that is associated with an existing snapshot or volume. However, you can associate a different CMK during a snapshot copy operation so that the resulting copied snapshot is encrypted by the new CMK.

**Creating New Empty Volumes with Encryption**

When you create a new, empty EBS volume, you can encrypt it by enabling encryption for the specific volume creation operation. If you enabled EBS encryption by default, the volume is automatically encrypted. By default, the volume is encrypted to your default key for EBS encryption. Alternatively, you can specify a different CMK for the specific volume creation operation. The volume is encrypted by the
time it is first available, so your data is always secured. For detailed procedures, see Creating an Amazon EBS Volume (p. 800).

By default, the CMK that you selected when creating a volume encrypts the snapshots that you make from the volume and the volumes that you restore from those encrypted snapshots. You cannot remove encryption from an encrypted volume or snapshot, which means that a volume restored from an encrypted snapshot, or a copy of an encrypted snapshot, is always encrypted.

Public snapshots of encrypted volumes are not supported, but you can share an encrypted snapshot with specific accounts. For detailed directions, see Sharing an Amazon EBS Snapshot (p. 842).

Encrypting Unencrypted Resources

Although there is no direct way to encrypt an existing unencrypted volume or snapshot, you can encrypt them by creating either a volume or a snapshot. If you enabled encryption by default, Amazon EBS encrypts the resulting new volume or snapshot using your default key for EBS encryption. Even if you have not enabled encryption by default, you can enable encryption when you create an individual volume or snapshot. Whether you enable encryption by default or in individual creation operations, you can override the default key for EBS encryption and select a customer managed CMK. For more information, see Creating an Amazon EBS Volume (p. 800) and Copying an Amazon EBS Snapshot (p. 837).

To encrypt the snapshot copy to a customer managed CMK, you must both enable encryption and specify the key, as shown in Copy an Unencrypted Snapshot (Encryption by Default Not Enabled) (p. 870).

You can also apply new encryption states when launching an instance from an EBS-backed AMI. This is because EBS-backed AMIs include snapshots of EBS volumes that can be encrypted as described. For more information, see Using Encryption with EBS-Backed AMIs (p. 67).

Encryption Scenarios

When you create an encrypted EBS resource, it is encrypted by your account's default key for EBS encryption unless you specify a different customer managed CMK in the volume creation parameters or the block device mapping for the AMI or instance. For more information, see Default Key for EBS Encryption (p. 867).

The following examples illustrate how you can manage the encryption state of your volumes and snapshots. For a full list of encryption cases, see the encryption outcomes table (p. 873).

Examples

- Restore an Unencrypted Volume (Encryption by Default Not Enabled) (p. 869)
- Restore an Unencrypted Volume (Encryption by Default Enabled) (p. 870)
- Copy an Unencrypted Snapshot (Encryption by Default Not Enabled) (p. 870)
- Copy an Unencrypted Snapshot (Encryption by Default Enabled) (p. 871)
- Re-Encrypt an Encrypted Volume (p. 871)
- Re-Encrypt an Encrypted Snapshot (p. 872)
- Migrate Data between Encrypted and Unencrypted Volumes (p. 872)
- Encryption Outcomes (p. 873)

Restore an Unencrypted Volume (Encryption by Default Not Enabled)

Without encryption by default enabled, a volume restored from an unencrypted snapshot is unencrypted by default. However, you can encrypt the resulting volume by setting the Encrypted parameter and, optionally, the KmsKeyId parameter. The following diagram illustrates the process.
If you leave out the KmsKeyId parameter, the resulting volume is encrypted using your default key for EBS encryption. You must specify a key ID to encrypt the volume to a different CMK.

For more information, see Restoring an Amazon EBS Volume from a Snapshot (p. 802).

**Restore an Unencrypted Volume (Encryption by Default Enabled)**

When you have enabled encryption by default, encryption is mandatory for volumes restored from unencrypted snapshots, and no encryption parameters are required for your default CMK to be used. The following diagram shows this simple default case:

If you want to encrypt the restored volume to a customer managed CMK, you must supply both the Encrypted and KmsKeyId parameters as shown in Restore an Unencrypted Volume (Encryption by Default Not Enabled) (p. 869).

**Copy an Unencrypted Snapshot (Encryption by Default Not Enabled)**

Without encryption by default enabled, a copy of an unencrypted snapshot is unencrypted by default. However, you can encrypt the resulting snapshot by setting the Encrypted parameter and, optionally, the KmsKeyId parameter. If you omit KmsKeyId, the resulting snapshot is encrypted by your default CMK. You must specify a key ID to encrypt the volume to a different CMK.

The following diagram illustrates the process.
Copy an Unencrypted Snapshot (Encryption by Default Enabled)

When you have enabled encryption by default, encryption is mandatory for copies of unencrypted snapshots, and no encryption parameters are required if your default CMK is used. The following diagram illustrates this default case:

Re-Encrypt an Encrypted Volume

When the CreateVolume action operates on an encrypted snapshot, you have the option of re-encrypting it with a different CMK. The following diagram illustrates the process. In this example, you own two CMKS, CMK A and CMK B. The source snapshot is encrypted by CMK A. During volume creation, with the key ID of CMK B specified as a parameter, the source data is automatically decrypted, then re-encrypted by CMK B.
Note
If you copy a snapshot and encrypt it to a new CMK, a complete (non-incremental) copy is always created, resulting in additional delay and storage costs.

For more information, see Restoring an Amazon EBS Volume from a Snapshot (p. 802).

Re-Encrypt an Encrypted Snapshot

The ability to encrypt a snapshot during copying allows you to apply a new CMK to an already-encrypted snapshot that you own. Volumes restored from the resulting copy are only accessible using the new CMK. The following diagram illustrates the process. In this example, you own two CMKs, CMK A and CMK B. The source snapshot is encrypted by CMK A. During copy, with the key ID of CMK B specified as a parameter, the source data is automatically re-encrypted by CMK B.

Note
If you copy a snapshot and encrypt it to a new CMK, a complete (non-incremental) copy is always created, resulting in additional delay and storage costs.

In a related scenario, you can choose to apply new encryption parameters to a copy of a snapshot that has been shared with you. By default, the copy is encrypted with a CMK shared by the snapshot’s owner. However, we recommend that you create a copy of the shared snapshot using a different CMK that you control. This protects your access to the volume if the original CMK is compromised, or if the owner revokes the CMK for any reason. For more information, see Encryption and Snapshot Copying (p. 839).

Migrate Data between Encrypted and Unencrypted Volumes

When you have access to both an encrypted and unencrypted volume, you can freely transfer data between them. EC2 carries out the encryption and decryption operations transparently.
For example, use the **robocopy** command to copy the data. In the following command, the source data is located in `D:\` and the destination volume is mounted at `E:\`.

```
PS C:\> robocopy D:\sourcefolder E:\destinationfolder /e /copyall /eta
```

We recommend using folders rather than copying an entire volume, as this avoids potential problems with hidden folders.

**Encryption Outcomes**

The following table describes the encryption outcome for each possible combination of settings.

<table>
<thead>
<tr>
<th>Is encryption enabled?</th>
<th>Is encryption by default enabled?</th>
<th>Source of volume</th>
<th>Default (no CMK specified)</th>
<th>Custom (CMK specified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>New (empty) volume</td>
<td>Unencrypted</td>
<td>N/A</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Unencrypted snapshot that you own</td>
<td>Unencrypted</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Encrypted snapshot that you own</td>
<td>Encrypted by same key</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Unencrypted snapshot that is shared with you</td>
<td>Unencrypted</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Encrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK*</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>New volume</td>
<td>Encrypted by default CMK</td>
<td>Encrypted by a specified CMK**</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Unencrypted snapshot that you own</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Encrypted snapshot that you own</td>
<td>Encrypted by same key</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Unencrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Encrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>New (empty) volume</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Unencrypted snapshot that you own</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Encrypted snapshot that you own</td>
<td>Encrypted by same key</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Unencrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Encrypted snapshot that is shared with you</td>
<td>Encrypted by default CMK</td>
<td></td>
</tr>
</tbody>
</table>
Is encryption enabled? | Is encryption by default enabled? | Source of volume | Default (no CMK specified) | Custom (CMK specified)
--- | --- | --- | --- | ---
Yes | Yes | New volume | Encrypted by default CMK | Encrypted by a specified CMK
Yes | Yes | Unencrypted snapshot that you own | Encrypted by default CMK | 
Yes | Yes | Encrypted snapshot that you own | Encrypted by same key | 
Yes | Yes | Unencrypted snapshot that is shared with you | Encrypted by default CMK | 
Yes | Yes | Encrypted snapshot that is shared with you | Encrypted by default CMK | 

* This is the default CMK used for EBS encryption for the AWS account and Region. By default this is a unique AWS managed CMK for EBS, or you can specify a customer managed CMK. For more information, see Default Key for EBS Encryption (p. 867).

** This is a customer managed CMK specified for the volume at launch time. This CMK is used instead of the default CMK for the AWS account and Region.

### Setting Encryption Defaults Using the API and CLI

You can manage encryption by default and the default customer master key (CMK) using the following API actions and CLI commands.

<table>
<thead>
<tr>
<th>API action</th>
<th>CLI command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DisableEbsEncryptionByDefault</td>
<td>disable-ebs-encryption-by-default</td>
<td>Disables encryption by default.</td>
</tr>
<tr>
<td>EnableEbsEncryptionByDefault</td>
<td>enable-ebs-encryption-by-default</td>
<td>Enables encryption by default.</td>
</tr>
<tr>
<td>GetEbsDefaultKmsKeyId</td>
<td>get-ebs-default-kms-key-id</td>
<td>Describes the default CMK.</td>
</tr>
<tr>
<td>GetEbsEncryptionByDefault</td>
<td>get-ebs-encryption-by-default</td>
<td>Indicates whether encryption by default is enabled.</td>
</tr>
<tr>
<td>ModifyEbsDefaultKmsKeyId</td>
<td>modify-ebs-default-kms-key-id</td>
<td>Changes the default CMK used to encrypt EBS volumes.</td>
</tr>
<tr>
<td>ResetEbsDefaultKmsKeyId</td>
<td>reset-ebs-default-kms-key-id</td>
<td>Resets the AWS managed default CMK as the default CMK used to encrypt EBS volumes.</td>
</tr>
</tbody>
</table>
Amazon EBS and NVMe on Windows Instances

EBS volumes are exposed as NVMe block devices on Nitro-based instances (p. 122).

Note
The EBS performance guarantees stated in Amazon EBS Product Details are valid regardless of the block-device interface.

Contents
- Install or Upgrade the NVMe Driver (p. 875)
- Identifying the EBS Device (p. 875)
- Working with NVMe EBS Volumes (p. 876)
- I/O Operation Timeout (p. 876)

Install or Upgrade the NVMe Driver

The AWS Windows AMIs for Windows Server 2008 R2 and later include the AWS NVMe driver. If you are not using the latest AWS Windows AMIs provided by Amazon, see Installing or Upgrading AWS NVMe Drivers (p. 446).

Identifying the EBS Device

EBS uses single-root I/O virtualization (SR-IOV) to provide volume attachments on Nitro-based instances using the NVMe specification. These devices rely on standard NVMe drivers on the operating system. These drivers typically discover attached devices by scanning the PCI bus during instance boot, and create device nodes based on the order in which the devices respond, not on how the devices are specified in the block device mapping.

Windows Server 2008 R2 and Later

You can also run the `ebsnvme-id` command to map the NVMe device disk number to an EBS volume ID and device name. By default, all EBS NVMe devices are enumerated. You can pass a disk number to enumerate information for a specific device. Ebsnvme-id is included in the latest AWS provided Windows Server AMIs located in C:\PROGRAMDATA\AMAZON\Tools.

You can also download `ebsnvme-id.zip` and extract the contents to your Amazon EC2 instance to get access to `ebsnvme-id.exe`.

```ps
disk Number: 0
Volume ID: vol-0d6d7ee96e471a7f
Device Name: sda1

disk Number: 1
Volume ID: vol-03a26248ff39b57cf
Device Name: xvdd

disk Number: 2
Volume ID: vol-038bd1c629a125e6
Device Name: xvde

disk Number: 3
Volume ID: vol-034f9d29ec064c89
Device Name: xvdb

disk Number: 4
Volume ID: vol-03e2db464b66f0a1
Device Name: xvdc
```

875
Working with NVMe EBS Volumes

The latest AWS Windows AMIs contain the AWS NVMe driver that is required by instance types that expose EBS volumes as NVMe block devices. However, if you resize your root volume on a Windows system, you must rescan the volume in order for this change to be reflected in the instance. If you launched your instance from a different AMI, it might not contain the required AWS NVMe driver. If your instance does not have the latest AWS NVMe driver, you must install it. For more information, see AWS NVMe Drivers for Windows Instances (p. 446).

I/O Operation Timeout

Most operating systems specify a timeout for I/O operations submitted to NVMe devices. On Windows systems, the default timeout is 60 seconds and the maximum is 255 seconds. You can modify the TimeoutValue disk class registry setting using the procedure described in Registry Entries for SCSI Miniport Drivers.

Amazon EBS–Optimized Instances

An Amazon EBS–optimized instance uses an optimized configuration stack and provides additional, dedicated capacity for Amazon EBS I/O. This optimization provides the best performance for your EBS volumes by minimizing contention between Amazon EBS I/O and other traffic from your instance.

EBS–optimized instances deliver dedicated bandwidth to Amazon EBS, with options between 425 MiB/s and 14,000 MiB/s, depending on the instance type you use. When attached to an EBS–optimized instance, General Purpose SSD (gp2) volumes are designed to deliver their baseline and burst performance 99% of the time, and Provisioned IOPS SSD (io1) volumes are designed to deliver their provisioned performance 99.9% of the time. Both Throughput Optimized HDD (st1) and Cold HDD (sc1) guarantee performance consistency of 90% of burst throughput 99% of the time. Non-compliant periods are approximately uniformly distributed, targeting 99% of expected total throughput each hour. For more information, see Amazon EBS Volume Types (p. 786).

Contents

- Supported Instance Types (p. 876)
- Enabling EBS Optimization at Launch (p. 885)
- Enable EBS Optimization for a Running Instance (p. 886)

Supported Instance Types

The following tables show which instance types support EBS optimization. They include the dedicated bandwidth to Amazon EBS, the typical maximum aggregate throughput that can be achieved on that connection with a streaming read workload and 128 KiB I/O size, and the maximum IOPS the instance can support if you are using a 16 KiB I/O size. Choose an EBS–optimized instance that provides more dedicated Amazon EBS throughput than your application needs; otherwise, the connection between Amazon EBS and Amazon EC2 can become a performance bottleneck.

EBS Optimized By Default

The following table lists the instance types that support EBS optimization and EBS optimization is enabled by default. There is no need to enable EBS optimization and no effect if you disable EBS optimization.
<table>
<thead>
<tr>
<th>Instance type</th>
<th>Maximum bandwidth (Mib/s)</th>
<th>Maximum throughput (MiB/s, 128 KiB I/O)</th>
<th>Maximum IOPS (16 KiB I/O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c4.large</td>
<td>500</td>
<td>62.5</td>
<td>4,000</td>
</tr>
<tr>
<td>c4.xlarge</td>
<td>750</td>
<td>93.75</td>
<td>6,000</td>
</tr>
<tr>
<td>c4.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>c4.4xlarge</td>
<td>2,000</td>
<td>250</td>
<td>16,000</td>
</tr>
<tr>
<td>c4.8xlarge</td>
<td>4,000</td>
<td>500</td>
<td>32,000</td>
</tr>
<tr>
<td>c5.large*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5.xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5.2xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5.9xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>c5.12xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>c5.18xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>c5.24xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>c5.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>c5d.large*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5d.xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5d.2xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5d.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5d.9xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>c5d.18xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>c5n.large*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5n.xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5n.2xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5n.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>c5n.9xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>c5n.18xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>c5n.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>d2.xlarge</td>
<td>750</td>
<td>93.75</td>
<td>6,000</td>
</tr>
<tr>
<td>d2.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>d2.4xlarge</td>
<td>2,000</td>
<td>250</td>
<td>16,000</td>
</tr>
<tr>
<td>Instance type</td>
<td>Maximum bandwidth (Mib/s)</td>
<td>Maximum throughput (MiB/s, 128 KiB I/O)</td>
<td>Maximum IOPS (16 KiB I/O)</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>d2.8xlarge</td>
<td>4,000</td>
<td>500</td>
<td>32,000</td>
</tr>
<tr>
<td>f1.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>f1.4xlarge</td>
<td>3,500</td>
<td>400</td>
<td>44,000</td>
</tr>
<tr>
<td>f1.16xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>75,000</td>
</tr>
<tr>
<td>g3s.xlarge</td>
<td>850</td>
<td>100</td>
<td>5,000</td>
</tr>
<tr>
<td>g3.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>g3.8xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>g3.16xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>g4dn.xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>10,000</td>
</tr>
<tr>
<td>g4dn.2xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>g4dn.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>g4dn.8xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>g4dn.12xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>g4dn.16xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>h1.2xlarge</td>
<td>1,750</td>
<td>218.75</td>
<td>12,000</td>
</tr>
<tr>
<td>h1.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>h1.8xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>h1.16xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>i3.large</td>
<td>425</td>
<td>53.13</td>
<td>3000</td>
</tr>
<tr>
<td>i3.xlarge</td>
<td>850</td>
<td>106.25</td>
<td>6000</td>
</tr>
<tr>
<td>i3.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>i3.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>16,000</td>
</tr>
<tr>
<td>i3.8xlarge</td>
<td>7,000</td>
<td>875</td>
<td>32,500</td>
</tr>
<tr>
<td>i3.16xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>65,000</td>
</tr>
<tr>
<td>i3.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>65,000</td>
</tr>
<tr>
<td>i3en.large*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>i3en.xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>i3en.2xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>i3en.3xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>i3en.6xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>Instance type</td>
<td>Maximum bandwidth (Mib/s)</td>
<td>Maximum throughput (MiB/s, 128 KiB I/O)</td>
<td>Maximum IOPS (16 KiB I/O)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>i3en.12xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>i3en.24xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>i3en.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>m4.large</td>
<td>450</td>
<td>56.25</td>
<td>3,600</td>
</tr>
<tr>
<td>m4.xlarge</td>
<td>750</td>
<td>93.75</td>
<td>6,000</td>
</tr>
<tr>
<td>m4.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>m4.4xlarge</td>
<td>2,000</td>
<td>250</td>
<td>16,000</td>
</tr>
<tr>
<td>m4.10xlarge</td>
<td>4,000</td>
<td>500</td>
<td>32,000</td>
</tr>
<tr>
<td>m4.16xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>65,000</td>
</tr>
<tr>
<td>m5.large*</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5.xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5.2xlarge*</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5.8xlarge</td>
<td>5,000</td>
<td>625</td>
<td>30,000</td>
</tr>
<tr>
<td>m5.12xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>m5.16xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>m5.24xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>m5.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>m5a.large*</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>m5a.xlarge*</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>m5a.2xlarge*</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>m5a.4xlarge</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>m5a.8xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>m5a.12xlarge</td>
<td>5,000</td>
<td>625</td>
<td>30,000</td>
</tr>
<tr>
<td>m5a.16xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>m5a.24xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>m5ad.large*</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>m5ad.xlarge*</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>m5ad.2xlarge*</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>m5ad.4xlarge</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>Instance type</td>
<td>Maximum bandwidth (Mib/s)</td>
<td>Maximum throughput (MiB/s, 128 KiB I/O)</td>
<td>Maximum IOPS (16 KiB I/O)</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------</td>
<td>----------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>m5ad.12xlarge</td>
<td>5,000</td>
<td>675</td>
<td>30,000</td>
</tr>
<tr>
<td>m5ad.24xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>m5d.large *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5d.xlarge *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5d.2xlarge *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5d.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>m5d.8xlarge</td>
<td>5,000</td>
<td>625</td>
<td>30,000</td>
</tr>
<tr>
<td>m5d.12xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>m5d.16xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>m5d.24xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>m5d.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>p2.xlarge</td>
<td>750</td>
<td>93.75</td>
<td>6,000</td>
</tr>
<tr>
<td>p2.8xlarge</td>
<td>5,000</td>
<td>625</td>
<td>32,500</td>
</tr>
<tr>
<td>p2.16xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>65,000</td>
</tr>
<tr>
<td>p3.2xlarge</td>
<td>1,750</td>
<td>218</td>
<td>10,000</td>
</tr>
<tr>
<td>p3.8xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>p3.16xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>p3dn.24xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>r4.large</td>
<td>425</td>
<td>53.13</td>
<td>3,000</td>
</tr>
<tr>
<td>r4.xlarge</td>
<td>850</td>
<td>106.25</td>
<td>6,000</td>
</tr>
<tr>
<td>r4.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>r4.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r4.8xlarge</td>
<td>7,000</td>
<td>875</td>
<td>37,500</td>
</tr>
<tr>
<td>r4.16xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>75,000</td>
</tr>
<tr>
<td>r5.large *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5.xlarge *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5.2xlarge *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5.8xlarge</td>
<td>5,000</td>
<td>625</td>
<td>30,000</td>
</tr>
<tr>
<td>r5.12xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>Instance type</td>
<td>Maximum bandwidth (Mib/s)</td>
<td>Maximum throughput (MiB/s, 128 KiB I/O)</td>
<td>Maximum IOPS (16 KiB I/O)</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------</td>
<td>-----------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>r5.16xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>r5.24xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>r5.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>r5a.large *</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5a.xlarge *</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5a.2xlarge *</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5a.4xlarge</td>
<td>2,120</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5a.8xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>32,000</td>
</tr>
<tr>
<td>r5a.12xlarge</td>
<td>5,000</td>
<td>625</td>
<td>30,000</td>
</tr>
<tr>
<td>r5a.16xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>r5a.24xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>r5ad.large *</td>
<td>2,210</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5ad.xlarge *</td>
<td>2,210</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5ad.2xlarge *</td>
<td>2,210</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5ad.4xlarge</td>
<td>2,210</td>
<td>265</td>
<td>16,000</td>
</tr>
<tr>
<td>r5ad.12xlarge</td>
<td>5,000</td>
<td>625</td>
<td>30,000</td>
</tr>
<tr>
<td>r5ad.24xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>r5d.large *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5d.xlarge *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5d.2xlarge *</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5d.4xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>18,750</td>
</tr>
<tr>
<td>r5d.8xlarge</td>
<td>5,000</td>
<td>625</td>
<td>30,000</td>
</tr>
<tr>
<td>r5d.12xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>r5d.16xlarge</td>
<td>10,000</td>
<td>1,250</td>
<td>60,000</td>
</tr>
<tr>
<td>r5d.24xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>r5d.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>t3.nano *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
<tr>
<td>t3.micro *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
<tr>
<td>t3.small *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
<tr>
<td>t3.medium *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
</tbody>
</table>
### EBS Optimization

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Maximum bandwidth (Mib/s)</th>
<th>Maximum throughput (MiB/s, 128 KiB I/O)</th>
<th>Maximum IOPS (16 KiB I/O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3.large *</td>
<td>2,048</td>
<td>256</td>
<td>15,700</td>
</tr>
<tr>
<td>t3.xlarge *</td>
<td>2,048</td>
<td>256</td>
<td>15,700</td>
</tr>
<tr>
<td>t3.2xlarge *</td>
<td>2,048</td>
<td>256</td>
<td>15,700</td>
</tr>
<tr>
<td>t3a.nano *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
<tr>
<td>t3a.micro *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
<tr>
<td>t3a.small *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
<tr>
<td>t3a.medium *</td>
<td>1,536</td>
<td>192</td>
<td>11,800</td>
</tr>
<tr>
<td>t3a.large *</td>
<td>2,048</td>
<td>256</td>
<td>15,700</td>
</tr>
<tr>
<td>t3a.xlarge *</td>
<td>2,048</td>
<td>256</td>
<td>15,700</td>
</tr>
<tr>
<td>t3a.2xlarge *</td>
<td>2,048</td>
<td>256</td>
<td>15,700</td>
</tr>
<tr>
<td>u-6tb1.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>u-9tb1.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>u-12tb1.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>x1.16xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>x1.32xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>x1e.xlarge</td>
<td>500</td>
<td>62.5</td>
<td>3,700</td>
</tr>
<tr>
<td>x1e.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>7,400</td>
</tr>
<tr>
<td>x1e.4xlarge</td>
<td>1,750</td>
<td>218.75</td>
<td>10,000</td>
</tr>
<tr>
<td>x1e.8xlarge</td>
<td>3,500</td>
<td>437.5</td>
<td>20,000</td>
</tr>
<tr>
<td>x1e.16xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>x1e.32xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>z1d.large *</td>
<td>2,333</td>
<td>291</td>
<td>13,333</td>
</tr>
<tr>
<td>z1d.xlarge *</td>
<td>2,333</td>
<td>291</td>
<td>13,333</td>
</tr>
<tr>
<td>z1d.2xlarge</td>
<td>2,333</td>
<td>292</td>
<td>13,333</td>
</tr>
<tr>
<td>z1d.3xlarge</td>
<td>3,500</td>
<td>438</td>
<td>20,000</td>
</tr>
<tr>
<td>z1d.6xlarge</td>
<td>7,000</td>
<td>875</td>
<td>40,000</td>
</tr>
<tr>
<td>z1d.12xlarge</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
<tr>
<td>z1d.metal</td>
<td>14,000</td>
<td>1,750</td>
<td>80,000</td>
</tr>
</tbody>
</table>

* These instance types can support maximum performance for 30 minutes at least once every 24 hours. For example, c5.large instances can deliver 437.5 MiB/s for 30 minutes at least once every 24 hours. If
you have a workload that requires sustained maximum performance for longer than 30 minutes, select an instance type according to baseline performance as shown in the following table:

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Baseline bandwidth (MiB/s)</th>
<th>Baseline throughput (MiB/s, 128 KiB I/O)</th>
<th>Baseline IOPS (16 KiB I/O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c5.large</td>
<td>525</td>
<td>65.625</td>
<td>4,000</td>
</tr>
<tr>
<td>c5.xlarge</td>
<td>800</td>
<td>100</td>
<td>6,000</td>
</tr>
<tr>
<td>c5.2xlarge</td>
<td>1,750</td>
<td>218.75</td>
<td>10,000</td>
</tr>
<tr>
<td>c5d.large</td>
<td>525</td>
<td>65.625</td>
<td>4,000</td>
</tr>
<tr>
<td>c5d.xlarge</td>
<td>800</td>
<td>100</td>
<td>6,000</td>
</tr>
<tr>
<td>c5d.2xlarge</td>
<td>1,750</td>
<td>218.75</td>
<td>10,000</td>
</tr>
<tr>
<td>c5n.large</td>
<td>525</td>
<td>65.625</td>
<td>4,000</td>
</tr>
<tr>
<td>c5n.xlarge</td>
<td>800</td>
<td>100</td>
<td>6,000</td>
</tr>
<tr>
<td>c5n.2xlarge</td>
<td>1,750</td>
<td>218.75</td>
<td>10,000</td>
</tr>
<tr>
<td>i3en.large</td>
<td>425</td>
<td>53.125</td>
<td>3,000</td>
</tr>
<tr>
<td>i3en.xlarge</td>
<td>850</td>
<td>106.25</td>
<td>6,000</td>
</tr>
<tr>
<td>i3en.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>i3en.3xlarge</td>
<td>2,800</td>
<td>350</td>
<td>15,000</td>
</tr>
<tr>
<td>m5.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>m5.xlarge</td>
<td>850</td>
<td>106.25</td>
<td>6,000</td>
</tr>
<tr>
<td>m5.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>m5a.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>m5a.xlarge</td>
<td>800</td>
<td>100</td>
<td>6,000</td>
</tr>
<tr>
<td>m5a.2xlarge</td>
<td>1,166</td>
<td>146</td>
<td>8,333</td>
</tr>
<tr>
<td>m5ad.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>m5ad.xlarge</td>
<td>800</td>
<td>100</td>
<td>6,000</td>
</tr>
<tr>
<td>m5ad.2xlarge</td>
<td>1,166</td>
<td>146</td>
<td>8,333</td>
</tr>
<tr>
<td>m5d.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>m5d.xlarge</td>
<td>850</td>
<td>106.25</td>
<td>6,000</td>
</tr>
<tr>
<td>m5d.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>r5.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>r5.xlarge</td>
<td>850</td>
<td>106.25</td>
<td>6,000</td>
</tr>
<tr>
<td>r5.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>r5a.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>Instance type</td>
<td>Baseline bandwidth (Mib/s)</td>
<td>Baseline throughput (MiB/s, 128 KiB I/O)</td>
<td>Baseline IOPS (16 KiB I/O)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>r5a.xlarge</td>
<td>800</td>
<td>100</td>
<td>6,000</td>
</tr>
<tr>
<td>r5a.2xlarge</td>
<td>1,166</td>
<td>146</td>
<td>8,333</td>
</tr>
<tr>
<td>r5ad.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>r5ad.xlarge</td>
<td>800</td>
<td>100</td>
<td>6,000</td>
</tr>
<tr>
<td>r5ad.2xlarge</td>
<td>1,166</td>
<td>146</td>
<td>8,333</td>
</tr>
<tr>
<td>r5d.large</td>
<td>480</td>
<td>60</td>
<td>3,600</td>
</tr>
<tr>
<td>r5d.xlarge</td>
<td>850</td>
<td>106.25</td>
<td>6,000</td>
</tr>
<tr>
<td>r5d.2xlarge</td>
<td>1,700</td>
<td>212.5</td>
<td>12,000</td>
</tr>
<tr>
<td>t3.nano</td>
<td>32</td>
<td>4</td>
<td>250</td>
</tr>
<tr>
<td>t3.micro</td>
<td>64</td>
<td>8</td>
<td>500</td>
</tr>
<tr>
<td>t3.small</td>
<td>128</td>
<td>16</td>
<td>1,000</td>
</tr>
<tr>
<td>t3.medium</td>
<td>256</td>
<td>32</td>
<td>2,000</td>
</tr>
<tr>
<td>t3.large</td>
<td>512</td>
<td>64</td>
<td>4,000</td>
</tr>
<tr>
<td>t3.xlarge</td>
<td>512</td>
<td>64</td>
<td>4,000</td>
</tr>
<tr>
<td>t3.2xlarge</td>
<td>512</td>
<td>64</td>
<td>4,000</td>
</tr>
<tr>
<td>t3a.nano</td>
<td>32</td>
<td>4</td>
<td>250</td>
</tr>
<tr>
<td>t3a.micro</td>
<td>64</td>
<td>8</td>
<td>500</td>
</tr>
<tr>
<td>t3a.small</td>
<td>128</td>
<td>16</td>
<td>1,000</td>
</tr>
<tr>
<td>t3a.medium</td>
<td>256</td>
<td>32</td>
<td>2,000</td>
</tr>
<tr>
<td>t3a.large</td>
<td>512</td>
<td>64</td>
<td>4,000</td>
</tr>
<tr>
<td>t3a.xlarge</td>
<td>512</td>
<td>64</td>
<td>4,000</td>
</tr>
<tr>
<td>t3a.2xlarge</td>
<td>512</td>
<td>64</td>
<td>4,000</td>
</tr>
<tr>
<td>z1d.large</td>
<td>583</td>
<td>73</td>
<td>3,333</td>
</tr>
<tr>
<td>z1d.xlarge</td>
<td>1,167</td>
<td>146</td>
<td>6,667</td>
</tr>
</tbody>
</table>

The EBSIOBalance% and EBSByteBalance% metrics can help you determine if your instances are sized correctly. You can view these metrics in the CloudWatch console and set an alarm that is triggered based on a threshold you specify. These metrics are expressed as a percentage. Instances with a consistently low balance percentage are candidates for upsizing. Instances where the balance percentage never drops below 100% are candidates for downsizing. For more information, see Monitoring Your Instances Using CloudWatch (p. 546).
EBS Optimization Supported

The following table lists the instance types that support EBS optimization but EBS optimization is not enabled by default. You can enable EBS optimization when you launch these instances or after they are running. Instances must have EBS optimization enabled to achieve the level of performance described. When you enable EBS optimization for an instance that is not EBS-optimized by default, you pay an additional low, hourly fee for the dedicated capacity. For pricing information, see EBS-optimized Instances on the Amazon EC2 Pricing page for On-Demand instances.

<table>
<thead>
<tr>
<th>Instance type</th>
<th>Maximum bandwidth (Mib/s)</th>
<th>Maximum throughput (MiB/s, 128 KiB I/O)</th>
<th>Maximum IOPS (16 KiB I/O)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1.xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>c3.xlarge</td>
<td>500</td>
<td>62.5</td>
<td>4,000</td>
</tr>
<tr>
<td>c3.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>c3.4xlarge</td>
<td>2,000</td>
<td>250</td>
<td>16,000</td>
</tr>
<tr>
<td>g2.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>i2.xlarge</td>
<td>500</td>
<td>62.5</td>
<td>4,000</td>
</tr>
<tr>
<td>i2.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>i2.4xlarge</td>
<td>2,000</td>
<td>250</td>
<td>16,000</td>
</tr>
<tr>
<td>m1.large</td>
<td>500</td>
<td>62.5</td>
<td>4,000</td>
</tr>
<tr>
<td>m1.xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>m2.2xlarge</td>
<td>500</td>
<td>62.5</td>
<td>4,000</td>
</tr>
<tr>
<td>m2.4xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>m3.xlarge</td>
<td>500</td>
<td>62.5</td>
<td>4,000</td>
</tr>
<tr>
<td>m3.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>r3.xlarge</td>
<td>500</td>
<td>62.5</td>
<td>4,000</td>
</tr>
<tr>
<td>r3.2xlarge</td>
<td>1,000</td>
<td>125</td>
<td>8,000</td>
</tr>
<tr>
<td>r3.4xlarge</td>
<td>2,000</td>
<td>250</td>
<td>16,000</td>
</tr>
</tbody>
</table>

The i2.8xlarge, c3.8xlarge, and r3.8xlarge instances do not have dedicated EBS bandwidth and therefore do not offer EBS optimization. On these instances, network traffic and Amazon EBS traffic share the same 10-gigabit network interface.

Enabling EBS Optimization at Launch

You can enable optimization for an instance by setting its attribute for EBS optimization.

To enable Amazon EBS optimization when launching an instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. Choose Launch Instance.
3. In **Step 1: Choose an Amazon Machine Image (AMI)**, select an AMI.

4. In **Step 2: Choose an Instance Type**, select an instance type that is listed as supporting Amazon EBS optimization.

5. In **Step 3: Configure Instance Details**, complete the fields that you need and choose **Launch as EBS-optimized instance**. If the instance type that you selected in the previous step doesn't support Amazon EBS optimization, this option is not present. If the instance type that you selected is Amazon EBS-optimized by default, this option is selected and you can't deselect it.

6. Follow the directions to complete the wizard and launch your instance.

**To enable EBS optimization when launching an instance using the command line**

You can use one of the following options with the corresponding command. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `--ebs-optimized` with `run-instances` (AWS CLI)
- `-EbsOptimized` with `New-EC2Instance` (AWS Tools for Windows PowerShell)

**Enable EBS Optimization for a Running Instance**

You can enable or disable optimization for a running instance by modifying its Amazon EBS-optimized instance attribute.

**To enable EBS optimization for a running instance using the console**

1. Open the Amazon EC2 console at `https://console.aws.amazon.com/ec2/`.
2. In the navigation pane, click **Instances**, and select the instance.
3. Click **Actions**, select **Instance State**, and then click **Stop**.

   **Warning**
   
   When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.

4. In the confirmation dialog box, click **Yes, Stop**. It can take a few minutes for the instance to stop.
5. With the instance still selected, click **Actions**, select **Instance Settings**, and then click **Change Instance Type**.
6. In the **Change Instance Type** dialog box, do one of the following:

   - If the instance type of your instance is Amazon EBS-optimized by default, **EBS-optimized** is selected and you can't change it. You can choose **Cancel**, because Amazon EBS optimization is already enabled for the instance.
   - If the instance type of your instance supports Amazon EBS optimization, choose **EBS-optimized**, **Apply**.
   - If the instance type of your instance does not support Amazon EBS optimization, you can't choose **EBS-optimized**. You can select an instance type from **Instance Type** that supports Amazon EBS optimization, and then choose **EBS-optimized**, **Apply**.
7. Choose **Actions**, **Instance State**, **Start**.

**To enable EBS optimization for a running instance using the command line**

You can use one of the following options with the corresponding command. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `--ebs-optimized` with `modify-instance-attribute` (AWS CLI)
Amazon EBS Volume Performance on Windows Instances

Several factors, including I/O characteristics and the configuration of your instances and volumes, can affect the performance of Amazon EBS. Customers who follow the guidance on our Amazon EBS and Amazon EC2 product detail pages typically achieve good performance out of the box. However, there are some cases where you may need to do some tuning in order to achieve peak performance on the platform. This topic discusses general best practices as well as performance tuning that is specific to certain use cases. We recommend that you tune performance with information from your actual workload, in addition to benchmarking, to determine your optimal configuration. After you learn the basics of working with EBS volumes, it’s a good idea to look at the I/O performance you require and at your options for increasing Amazon EBS performance to meet those requirements.

**Note**

AWS updates to the performance of EBS volume types may not immediately take effect on your existing volumes. To see full performance on an older volume, you may first need to perform a `ModifyVolume` action on it. For more information, see Modifying the Size, IOPS, or Type of an EBS Volume on Windows.

Contents

- Amazon EBS Performance Tips (p. 887)
- Amazon EC2 Instance Configuration (p. 888)
- I/O Characteristics and Monitoring (p. 889)
- Initializing Amazon EBS Volumes (p. 891)
- RAID Configuration on Windows (p. 894)
- Benchmark EBS Volumes (p. 898)

Amazon EBS Performance Tips

These tips represent best practices for getting optimal performance from your EBS volumes in a variety of user scenarios.

**Use EBS-Optimized Instances**

On instances without support for EBS-optimized throughput, network traffic can contend with traffic between your instance and your EBS volumes; on EBS-optimized instances, the two types of traffic are kept separate. Some EBS-optimized instance configurations incur an extra cost (such as C3, R3, and M3), while others are always EBS-optimized at no extra cost (such as M4, C4, C5, and D2). For more information, see Amazon EC2 Instance Configuration (p. 888).

**Understand How Performance is Calculated**

When you measure the performance of your EBS volumes, it is important to understand the units of measure involved and how performance is calculated. For more information, see I/O Characteristics and Monitoring (p. 889).

**Understand Your Workload**

There is a relationship between the maximum performance of your EBS volumes, the size and number of I/O operations, and the time it takes for each action to complete. Each of these factors (performance, I/O, and latency) affects the others, and different applications are more sensitive to one factor or another.
Be Aware of the Performance Penalty When Initializing Volumes from Snapshots

There is a significant increase in latency when you first access each block of data on a new EBS volume that was restored from a snapshot. You can avoid this performance hit by accessing each block prior to putting the volume into production. This process is called initialization (formerly known as pre-warming). For more information, see Initializing Amazon EBS Volumes (p. 891).

Factors That Can Degrade HDD Performance

When you create a snapshot of a Throughput Optimized HDD (st1) or Cold HDD (sc1) volume, performance may drop as far as the volume's baseline value while the snapshot is in progress. This behavior is specific to these volume types. Other factors that can limit performance include driving more throughput than the instance can support, the performance penalty encountered while initializing volumes restored from a snapshot, and excessive amounts of small, random I/O on the volume. For more information about calculating throughput for HDD volumes, see Amazon EBS Volume Types (p. 786).

Your performance can also be impacted if your application isn't sending enough I/O requests. This can be monitored by looking at your volume's queue length and I/O size. The queue length is the number of pending I/O requests from your application to your volume. For maximum consistency, HDD-backed volumes must maintain a queue length (rounded to the nearest whole number) of 4 or more when performing 1 MiB sequential I/O. For more information about ensuring consistent performance of your volumes, see I/O Characteristics and Monitoring (p. 889).

Use RAID 0 to Maximize Utilization of Instance Resources

Some instance types can drive more I/O throughput than what you can provision for a single EBS volume. You can join multiple gp2, io1, st1, or sc1 volumes together in a RAID 0 configuration to use the available bandwidth for these instances. For more information, see RAID Configuration on Windows (p. 894).

Track Performance Using Amazon CloudWatch

Amazon Web Services provides performance metrics for Amazon EBS that you can analyze and view with Amazon CloudWatch and status checks that you can use to monitor the health of your volumes. For more information, see Monitoring the Status of Your Volumes (p. 808).

Amazon EC2 Instance Configuration

When you plan and configure EBS volumes for your application, it is important to consider the configuration of the instances that you will attach the volumes to. In order to get the most performance out of your EBS volumes, you should attach them to an instance with enough bandwidth to support your volumes, such as an EBS-optimized instance or an instance with 10 gigabit network connectivity. This is especially important when you stripe multiple volumes together in a RAID configuration.

Use EBS-Optimized or 10 Gigabit Network Instances

Any performance-sensitive workloads that require minimal variability and dedicated Amazon EC2 to Amazon EBS traffic, such as production databases or business applications, should use volumes that are attached to an EBS-optimized instance or an instance with 10 gigabit network connectivity. EC2 instances that do not meet this criteria offer no guarantee of network resources. The only way to ensure sustained reliable network bandwidth between your EC2 instance and your EBS volumes is to launch the EC2 instance as EBS-optimized or choose an instance type with 10 gigabit network connectivity. To see which instance types include 10 gigabit network connectivity, see Amazon EC2 Instance Types. For information about configuring EBS-optimized instances, see Amazon EBS–Optimized Instances (p. 876).

Choose an EC2 Instance with Enough Bandwidth

Launching an instance that is EBS-optimized provides you with a dedicated connection between your EC2 instance and your EBS volume. However, it is still possible to provision EBS volumes that exceed...
the available bandwidth for certain instance types, especially when multiple volumes are striped in a RAID configuration. For information about the instance types available to be launched as EBS-optimized, the dedicated throughput to these instance types, the dedicated bandwidth to Amazon EBS, the maximum amount of IOPS the instance can support if you are using a 16 KiB I/O size, and the approximate I/O bandwidth available on that connection, see Supported Instance Types (p. 876).

Be sure to choose an EBS-optimized instance that provides more dedicated EBS throughput than your application needs; otherwise, the Amazon EBS to Amazon EC2 connection becomes a performance bottleneck.

Note that some instances with 10-gigabit network interfaces do not offer EBS-optimization, and therefore do not have dedicated EBS bandwidth available. However, you can use all of that bandwidth for traffic to Amazon EBS if your application isn’t pushing other network traffic that contends with Amazon EBS. Some 10-gigabit network instances offer dedicated Amazon EBS bandwidth in addition to a 10-gigabit interface which is used exclusively for network traffic.

If an instance type has a maximum 16 KB IOPS value of 4,000, that value is an absolute best-case scenario and is not guaranteed unless the instance is launched as EBS-optimized. To consistently achieve the best performance, you must launch instances as EBS-optimized. However, if you attach a 4,000 IOPS \( \text{io1} \) volume to an EBS-optimized instance with a 16 KB IOPS value of 4,000, the Amazon EC2 to Amazon EBS connection bandwidth limit prevents this volume from providing the 500 MiB/s maximum aggregate throughput available to it. In this case, we must use an EBS-optimized EC2 instance that supports at least 500 MiB/s of throughput.

Volumes of type General Purpose SSD (\( \text{gp2} \)) have a throughput limit between 128 MiB/s and 250 MiB/s per volume (depending on volume size), which pairs well with a 1,000 MiB/s EBS-optimized connection. Instance types that offer more than 1,000 MiB/s of throughput to Amazon EBS can use more than one \( \text{gp2} \) volume to take advantage of the available throughput. Volumes of type Provisioned IOPS SSD (\( \text{io1} \)) have a throughput limit range of 256 KiB for each IOPS provisioned, up to a maximum of 1,000 MiB/s (at 64,000 IOPS). For more information, see Amazon EBS Volume Types (p. 786).

Note
These performance values for \( \text{io1} \) are guaranteed only for volumes attached to Nitro-based instances. For other instances, AWS guarantees performance up to 500 MiB/s and 32,000 IOPS per volume. For more information, see Amazon EBS Volume Types (p. 786).

Instance types with 10 gigabit network connectivity support up to 800 MiB/s of throughput and 48,000 16K IOPS for unencrypted Amazon EBS volumes and up to 25,000 16K IOPS for encrypted Amazon EBS volumes. Because the maximum \( \text{io1} \) value for EBS volumes is 64,000 for \( \text{io1} \) volumes and 16,000 for \( \text{gp2} \) volumes, you can use several EBS volumes simultaneously to reach the level of I/O performance available to these instance types. For more information about which instance types include 10 gigabit network connectivity, see Amazon EC2 Instance Types.

You should use EBS-optimized instances when available to get the full performance benefits of Amazon EBS \( \text{gp2} \) and \( \text{io1} \) volumes. For more information, see Amazon EBS–Optimized Instances (p. 876).

I/O Characteristics and Monitoring

On a given volume configuration, certain I/O characteristics drive the performance behavior for your EBS volumes. SSD-backed volumes—General Purpose SSD (\( \text{gp2} \)) and Provisioned IOPS SSD (\( \text{io1} \))—deliver consistent performance whether an I/O operation is random or sequential. HDD-backed volumes—Throughput Optimized HDD (\( \text{st1} \)) and Cold HDD (\( \text{sc1} \))—deliver optimal performance only when I/O operations are large and sequential. To understand how SSD and HDD volumes will perform in your application, it is important to know the connection between demand on the volume, the quantity of IOPS available to it, the time it takes for an I/O operation to complete, and the volume's throughput limits.

IOPS
IOPS are a unit of measure representing input/output operations per second. The operations are measured in KiB, and the underlying drive technology determines the maximum amount of data that a volume type counts as a single I/O. I/O size is capped at 256 KiB for SSD volumes and 1,024 KiB for HDD volumes because SSD volumes handle small or random I/O much more efficiently than HDD volumes.

When small I/O operations are physically contiguous, Amazon EBS attempts to merge them into a single I/O operation up to the maximum size. For example, for SSD volumes, a single 1,024 KiB I/O operation counts as 4 operations (1,024 ÷ 256 = 4), while 8 contiguous I/O operations at 32 KiB each count as 1 operation (8×32=256). However, 8 random I/O operations at 32 KiB each count as 8 operations. Each I/O operation under 32 KiB counts as 1 operation.

Similarly, for HDD-backed volumes, both a single 1,024 KiB I/O operation and 8 sequential 128 KiB operations would count as one operation. However, 8 random 128 KiB I/O operations would count as 8 operations.

Consequently, when you create an SSD-backed volume supporting 3,000 IOPS (either by provisioning an io1 volume at 3,000 IOPS or by sizing a gp2 volume at 1000 GiB), and you attach it to an EBS-optimized instance that can provide sufficient bandwidth, you can transfer up to 3,000 I/Os of data per second, with throughput determined by I/O size.

Volume Queue Length and Latency

The volume queue length is the number of pending I/O requests for a device. Latency is the true end-to-end client time of an I/O operation, in other words, the time elapsed between sending an I/O to EBS and receiving an acknowledgement from EBS that the I/O read or write is complete. Queue length must be correctly calibrated with I/O size and latency to avoid creating bottlenecks either on the guest operating system or on the network link to EBS.

Optimal queue length varies for each workload, depending on your particular application’s sensitivity to IOPS and latency. If your workload is not delivering enough I/O requests to fully use the performance available to your EBS volume, then your volume might not deliver the IOPS or throughput that you have provisioned.

Transaction-intensive applications are sensitive to increased I/O latency and are well-suited for SSD-backed io1 and gp2 volumes. You can maintain high IOPS while keeping latency down by maintaining a low queue length and a high number of IOPS available to the volume. Consistently driving more IOPS to a volume than it has available can cause increased I/O latency.

Throughput-intensive applications are less sensitive to increased I/O latency, and are well-suited for HDD-backed st1 and sc1 volumes. You can maintain high throughput to HDD-backed volumes by maintaining a high queue length when performing large, sequential I/O.

I/O size and volume throughput limits

For SSD-backed volumes, if your I/O size is very large, you may experience a smaller number of IOPS than you provisioned because you are hitting the throughput limit of the volume. For example, a gp2 volume under 1000 GiB with burst credits available has an IOPS limit of 3,000 and a volume throughput limit of 250 MiB/s. If you are using a 256 KiB I/O size, your volume reaches its throughput limit at 1000 IOPS (1000 x 256 KiB = 250 MiB). For smaller I/O sizes (such as 16 KiB), this same volume can sustain 3,000 IOPS because the throughput is well below 250 MiB/s. (These examples assume that your volume’s I/O is not hitting the throughput limits of the instance.) For more information about the throughput limits for each EBS volume type, see Amazon EBS Volume Types (p. 786).

For smaller I/O operations, you may see a higher-than-provisioned IOPS value as measured from inside your instance. This happens when the instance operating system merges small I/O operations into a larger operation before passing them to Amazon EBS.

If your workload uses sequential I/Os on HDD-backed st1 and sc1 volumes, you may experience a higher than expected number of IOPS as measured from inside your instance. This happens when the instance operating system merges sequential I/Os and counts them in 1,024 KiB-sized units. If your
workload uses small or random I/Os, you may experience a lower throughput than you expect. This is
because we count each random, non-sequential I/O toward the total IOPS count, which can cause you to
hit the volume's IOPS limit sooner than expected.

Whatever your EBS volume type, if you are not experiencing the IOPS or throughput you expect in your
configuration, ensure that your EC2 instance bandwidth is not the limiting factor. You should always use
a current-generation, EBS-optimized instance (or one that includes 10 Gb/s network connectivity) for
optimal performance. For more information, see Amazon EC2 Instance Configuration (p. 888). Another
possible cause for not experiencing the expected IOPS is that you are not driving enough I/O to the EBS
volumes.

Monitor I/O Characteristics with CloudWatch

You can monitor these I/O characteristics with each volume's CloudWatch metrics (p. 808). Important
metrics to consider include:

- BurstBalance
- VolumeReadBytes
- VolumeWriteBytes
- VolumeReadOps
- VolumeWriteOps
- VolumeQueueLength

**BurstBalance** displays the burst bucket balance for gp2, st1, and sc1 volumes as a percentage of
the remaining balance. When your burst bucket is depleted, volume I/O (for gp2 volumes) or volume
throughput (for st1 and sc1 volumes) is throttled to the baseline. Check the **BurstBalance** value to
determine whether your volume is being throttled for this reason.

HDD-backed st1 and sc1 volumes are designed to perform best with workloads that take
advantage of the 1,024 KiB maximum I/O size. To determine your volume's average I/O size, divide
**VolumeWriteBytes** by **VolumeWriteOps**. The same calculation applies to read operations. If average
I/O size is below 64 KiB, increasing the size of the I/O operations sent to an st1 or sc1 volume should
improve performance.

**Note**

If average I/O size is at or near 44 KiB, you may be using an instance or kernel without support
for indirect descriptors. Any Linux kernel 3.8 and above has this support, as well as any current-
generation instance.

If your I/O latency is higher than you require, check **VolumeQueueLength** to make sure your application
is not trying to drive more IOPS than you have provisioned. If your application requires a greater number
of IOPS than your volume can provide, you should consider using a larger gp2 volume with a higher base
performance level or an io1 volume with more provisioned IOPS to achieve faster latencies.

For more information about Amazon EBS I/O characteristics, see the Amazon EBS: Designing for
Performance re:Invent presentation on this topic.

**Initializing Amazon EBS Volumes**

New EBS volumes receive their maximum performance the moment that they are available and do not
require initialization (formerly known as pre-warming).

However, storage blocks on volumes that were restored from snapshots must be pulled down from
Amazon S3 and written to the volume before they can be accessed. This preliminary action takes time
and can significantly increase the latency of I/O operations. For most applications, amortizing this cost
over the lifetime of the volume is acceptable. The volume's performance is restored after all of the blocks
have been downloaded and written to the volume. To avoid this initial performance hit in a production
environment, you can perform an initialization to read all of the blocks on the volume before you use it.
Important

While initializing io1 volumes that were restored from snapshots, the performance of the volume may drop below 50 percent of its expected level, which causes the volume to display a warning state in the I/O Performance status check. This is expected, and you can ignore the warning state on io1 volumes while you are initializing them. For more information, see EBS Volume Status Checks (p. 809).

Initializing Amazon EBS Volumes on Windows

New EBS volumes receive their maximum performance the moment that they are available and do not require initialization (formerly known as pre-warming). For volumes that have been restored from snapshots, use `dd` or `fio` for Windows to read from all of the blocks on a volume. All existing data on the volume will be preserved.

Before using either tool, gather information about the disks on your system as follows:

1. Use the `wmic` command to list the available disks on your system:

   `wmic diskdrive get size,deviceid`

   The following is example output:

<table>
<thead>
<tr>
<th>DeviceID</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>.\PHYSICALDRIVE2</td>
<td>80517265920</td>
</tr>
<tr>
<td>.\PHYSICALDRIVE1</td>
<td>80517265920</td>
</tr>
<tr>
<td>.\PHYSICALDRIVE0</td>
<td>128849011200</td>
</tr>
<tr>
<td>.\PHYSICALDRIVE3</td>
<td>107372805120</td>
</tr>
</tbody>
</table>

2. Identify the disk to initialize using `dd` or `fio`. The C: drive is on `\.\PHYSICALDRIVE0`. You can use the `diskmgmt.msc` utility to compare drive letters to disk drive numbers if you are not sure which drive number to use.

Using `dd`

Complete the following procedures to install and use `dd` to initialize a volume.

Note

This step may take several minutes up to several hours, depending on your EC2 instance bandwidth, the IOPS provisioned for the volume, and the size of the volume.

Install `dd` for Windows

The `dd` for Windows program provides a similar experience to the `dd` program that is commonly available for Linux and Unix systems, and it allows you to initialize Amazon EBS volumes that have been restored from snapshots. At the time of this writing, the most recent beta version contains the `/dev/null` virtual device that is required to initialize volumes restored from snapshots. Full documentation for the program is available at http://www.chrysocome.net/dd.

1. Download the most recent binary version of `dd` for Windows from http://www.chrysocome.net/dd. You must use version 0.6 beta 3 or newer to initialize restored volumes.
2. (Optional) Create a folder for command line utilities that is easy to locate and remember, such as `C:\bin`. If you already have a designated folder for command line utilities, you can use that folder instead in the following step.
3. Unzip the binary package and copy the `dd.exe` file to your command line utilities folder (for example, `C:\bin`).
4. Add the command line utilities folder to your Path environment variable so you can execute the programs in that folder from anywhere.
Important
The following steps don't update the environment variables in your current command prompt windows. The command prompt windows that you open after you complete these steps will contain the updates. This is why it's necessary for you to open a new command prompt window to verify that your environment is set up properly.

a. Choose Start, open the context (right-click) menu for Computer, and then choose Properties.
b. Choose Advanced system settings, Environment Variables.
c. For System Variables, select the variable Path and choose Edit.
d. For Variable value, append a semicolon and the location of your command line utility folder (;C:\bin) to the end of the existing value.
e. Choose OK to close the Edit System Variable window.

Initialize a volume using dd for Windows
1. Execute the following command to read all blocks on the specified device (and send the output to the /dev/null virtual device). This command safely initializes your existing data.

   Important
   Incorrect use of dd can easily destroy a volume's data. Be sure to follow precisely the example command below. Only the if=\\.\\PHYSICALDRIVEn parameter will vary depending on the name of the device you are reading.

   ```
   dd if=\\.\\PHYSICALDRIVEn of=/dev/null bs=1M --progress --size
   ```

   Note
   You may see an error if dd attempts to read beyond the end of the volume. This can be safely ignored.

2. When the operation completes, you are ready to use your new volume. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804).

Using fio
Complete the following procedures to install and use fio to initialize a volume.

Install fio for Windows
The fio for Windows program provides a similar experience to the fio program that is commonly available for Linux and Unix systems, and it allows you to initialize Amazon EBS volumes that have been restored from snapshots. Full documentation for the program is available at https://github.com/axboe/fio.

1. Download the fio MSI installer (select the latest x86 or x64 build, then select Artifacts).
2. Install fio.

Initialize a volume using fio for Windows
1. Run a command similar to the following to initialize a volume:

   ```
   fio --filename=\\.\\PHYSICALDRIVEn --rw=read --bs=128k --iodepth=32 --direct=1 --name=volume-initialize
   ```

2. When the operation completes, you are ready to use your new volume. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804).
RAID Configuration on Windows

With Amazon EBS, you can use any of the standard RAID configurations that you can use with a traditional bare metal server, as long as that particular RAID configuration is supported by the operating system for your instance. This is because all RAID is accomplished at the software level. For greater I/O performance than you can achieve with a single volume, RAID 0 can stripe multiple volumes together; for on-instance redundancy, RAID 1 can mirror two volumes together.

Amazon EBS volume data is replicated across multiple servers in an Availability Zone to prevent the loss of data from the failure of any single component. This replication makes Amazon EBS volumes ten times more reliable than typical commodity disk drives. For more information, see Amazon EBS Availability and Durability in the Amazon EBS product detail pages.

**Note**
You should avoid booting from a RAID volume. If one of the devices fails, you may be unable to boot the operating system.

If you need to create a RAID array on a Linux instance, see RAID Configuration on Linux in the Amazon EC2 User Guide for Linux Instances.

Contents

- RAID Configuration Options (p. 894)
- Creating a RAID Array on Windows (p. 895)
- Creating Snapshots of Volumes in a RAID Array (p. 898)

RAID Configuration Options

The following table compares the common RAID 0 and RAID 1 options.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Use</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAID 0</td>
<td>When I/O performance is more important than fault tolerance; for example, as in a heavily used database (where data replication is already set up separately).</td>
<td>I/O is distributed across the volumes in a stripe. If you add a volume, you get the straight addition of throughput.</td>
<td>Performance of the stripe is limited to the worst performing volume in the set. Loss of a single volume results in a complete data loss for the array.</td>
</tr>
<tr>
<td>RAID 1</td>
<td>When fault tolerance is more important than I/O performance; for example, as in a critical application.</td>
<td>Safer from the standpoint of data durability.</td>
<td>Does not provide a write performance improvement; requires more Amazon EC2 to Amazon EBS bandwidth than non-RAID configurations because the data is written to multiple volumes simultaneously.</td>
</tr>
</tbody>
</table>

**Important**
RAID 5 and RAID 6 are not recommended for Amazon EBS because the parity write operations of these RAID modes consume some of the IOPS available to your volumes. Depending on the configuration of your RAID array, these RAID modes provide 20-30% fewer usable IOPS than a RAID 0 configuration. Increased cost is a factor with these RAID modes as well; when using identical volume sizes and speeds, a 2-volume RAID 0 array can outperform a 4-volume RAID 6 array that costs twice as much.
Creating a RAID 0 array allows you to achieve a higher level of performance for a file system than you can provision on a single Amazon EBS volume. A RAID 1 array offers a "mirror" of your data for extra redundancy. Before you perform this procedure, you need to decide how large your RAID array should be and how many IOPS you want to provision.

The resulting size of a RAID 0 array is the sum of the sizes of the volumes within it, and the bandwidth is the sum of the available bandwidth of the volumes within it. The resulting size and bandwidth of a RAID 1 array is equal to the size and bandwidth of the volumes in the array. For example, two 500 GiB Amazon EBS io1 volumes with 4,000 provisioned IOPS each will create a 1000 GiB RAID 0 array with an available bandwidth of 8,000 IOPS and 1,000 MiB/s of throughput or a 500 GiB RAID 1 array with an available bandwidth of 4,000 IOPS and 500 MiB/s of throughput.

This documentation provides basic RAID setup examples. For more information about RAID configuration, performance, and recovery, see the Linux RAID Wiki at https://raid.wiki.kernel.org/index.php/Linux_Raid.

Creating a RAID Array on Windows

Use the following procedure to create the RAID array. Note that you can get directions for Linux instances from Creating a RAID Array on Linux in the Amazon EC2 User Guide for Linux Instances.

To create a RAID array on Windows

1. Create the Amazon EBS volumes for your array. For more information, see Creating an Amazon EBS Volume (p. 800).

   **Important**
   Create volumes with identical size and IOPS performance values for your array. Make sure you do not create an array that exceeds the available bandwidth of your EC2 instance. For more information, see Amazon EC2 Instance Configuration (p. 888).

2. Attach the Amazon EBS volumes to the instance that you want to host the array. For more information, see Attaching an Amazon EBS Volume to an Instance (p. 803).

3. Connect to your Windows instance. For more information, see Connecting to Your Windows Instance (p. 379).

4. Open a command prompt and type the `diskpart` command.

   ```
   diskpart
   Microsoft DiskPart version 6.1.7601
   Copyright (C) 1999-2008 Microsoft Corporation.
   On computer: WIN-BM6QPPL51CO
   ```

5. At the DISKPART prompt, list the available disks with the following command.

   ```
   DISKPART> list disk
   Disk ###  Status         Size     Free     Dyn  Gpt
            --------  -------------  -------  -------  ---  ---
   Disk 0    Online           30 GB      0 B
   Disk 1    Online            8 GB      0 B
   Disk 2    Online            8 GB      0 B
   Disk 3    Online            8 GB      0 B
   Disk 4    Online            8 GB      0 B
   Disk 5    Online          419 GB      0 B
   Disk 6    Online          419 GB      0 B
   ```

Identify the disks you want to use in your array and take note of their disk numbers.
6. Each disk you want to use in your array must be an online dynamic disk that does not contain any existing volumes. Use the following steps to convert basic disks to dynamic disks and to delete any existing volumes.

   a. Select a disk you want to use in your array with the following command, substituting $n$ with your disk number.

   ```
   DISKPART> select disk n
   Disk n is now the selected disk.
   ```

   b. If the selected disk is listed as Offline, bring it online by running the `online disk` command.

   c. If the selected disk does not have an asterisk in the `Dyn` column in the previous `list disk` command output, you need to convert it to a dynamic disk.

   ```
   DISKPART> convert dynamic
   ```

   **Note**
   If you receive an error that the disk is write protected, you can clear the read-only flag with the `ATTRIBUTE DISK CLEAR READONLY` command and then try the dynamic disk conversion again.

   d. Use the `detail disk` command to check for existing volumes on the selected disk.

   ```
   DISKPART> detail disk
   ```

   Note any volume numbers on the disk. In this example, the volume number is 2. If there are no volumes, you can skip the next step.

   e. (Only required if volumes were identified in the previous step) Select and delete any existing volumes on the disk that you identified in the previous step.

   **Warning**
   This destroys any existing data on the volume.

   i. Select the volume, substituting $n$ with your volume number.

   ```
   DISKPART> select volume n
   Volume n is the selected volume.
   ```

   ii. Delete the volume.
iii. Repeat these substeps for each volume you need to delete on the selected disk.

f. Repeat Step 6 (p. 896) for each disk you want to use in your array.

7. Verify that the disks you want to use are now dynamic.

<table>
<thead>
<tr>
<th>Disk ###</th>
<th>Status</th>
<th>Size</th>
<th>Free</th>
<th>Dyn</th>
<th>Gpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk 0</td>
<td>Online</td>
<td>30 GB</td>
<td>0 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disk 1</td>
<td>Online</td>
<td>8 GB</td>
<td>0 B</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Disk 2</td>
<td>Online</td>
<td>8 GB</td>
<td>0 B</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Disk 3</td>
<td>Online</td>
<td>8 GB</td>
<td>0 B</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>* Disk 4</td>
<td>Online</td>
<td>8 GB</td>
<td>0 B</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Disk 5</td>
<td>Online</td>
<td>419 GB</td>
<td>0 B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disk 6</td>
<td>Online</td>
<td>419 GB</td>
<td>0 B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Create your raid array. On Windows, a RAID 0 volume is referred to as a striped volume and a RAID 1 volume is referred to as a mirrored volume.

(Striped volumes only) To create a striped volume array on disks 1 and 2, use the following command (note the stripe option to stripe the array):

```
DISKPART> create volume stripe disk=1,2
```

DiskPart successfully created the volume.

(Mirrored volumes only) To create a mirrored volume array on disks 3 and 4, use the following command (note the mirror option to mirror the array):

```
DISKPART> create volume mirror disk=3,4
```

DiskPart successfully created the volume.


<table>
<thead>
<tr>
<th>Volume ###</th>
<th>Ltr</th>
<th>Label</th>
<th>Fs</th>
<th>Type</th>
<th>Size</th>
<th>Status</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume 0</td>
<td>C</td>
<td>NTFS</td>
<td>Partition</td>
<td>29 GB</td>
<td>Healthy</td>
<td>System</td>
<td></td>
</tr>
<tr>
<td>* Volume 1</td>
<td></td>
<td>RAW</td>
<td>Mirror</td>
<td>8190 MB</td>
<td>Healthy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume 2</td>
<td></td>
<td>RAW</td>
<td>Stripe</td>
<td>15 GB</td>
<td>Healthy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume 5</td>
<td>Z</td>
<td>Temporary</td>
<td>S</td>
<td>NTFS</td>
<td>Partition</td>
<td>419 GB</td>
<td>Healthy</td>
</tr>
<tr>
<td>Volume 6</td>
<td>Y</td>
<td>Temporary</td>
<td>S</td>
<td>NTFS</td>
<td>Partition</td>
<td>419 GB</td>
<td>Healthy</td>
</tr>
</tbody>
</table>

Note that for this example the Type column lists a Mirror volume and a Stripe volume.

10. Select and format your volume so that you can begin using it.

a. Select the volume you want to format, substituting n with your volume number.

```
DISKPART> select volume n
```

Volume n is the selected volume.
b. Format the volume.

   **Note**
   To perform a full format, omit the quick option.

   ```plaintext
   DISKPART> format quick recommended label="My new volume"
   100 percent completed
   DiskPart successfully formatted the volume.
   ```

   c. Assign an available drive letter to your volume.

   ```plaintext
   DISKPART> assign letter f
   DiskPart successfully assigned the drive letter or mount point.
   ```

   Your new volume is now ready to use.

**Creating Snapshots of Volumes in a RAID Array**

If you want to back up the data on the EBS volumes in a RAID array using snapshots, you must ensure that the snapshots are consistent. This is because the snapshots of these volumes are created independently. To restore EBS volumes in a RAID array from snapshots that are out of sync would degrade the integrity of the array.

To create a consistent set of snapshots for your RAID array, use **EBS multi-volume snapshots**. Multi-volume snapshots allow you to take point-in-time, data coordinated, and crash-consistent snapshots across multiple EBS volumes attached to an EC2 instance. You do not have to stop your instance to coordinate between volumes to ensure consistency because snapshots are automatically taken across multiple EBS volumes. For more information, see the steps for creating multi-volume snapshots under **Creating Amazon EBS Snapshots**.

**Benchmark EBS Volumes**

You can test the performance of Amazon EBS volumes by simulating I/O workloads. The process is as follows:

1. Launch an EBS-optimized instance.
2. Create new EBS volumes.
3. Attach the volumes to your EBS-optimized instance.
4. Configure and mount the block device.
5. Install a tool to benchmark I/O performance.
7. Delete your volumes and terminate your instance so that you don't continue to incur charges.

   **Important**
   Some of the procedures result in the destruction of existing data on the EBS volumes you benchmark. The benchmarking procedures are intended for use on volumes specially created for testing purposes, not production volumes.
Set Up Your Instance

To get optimal performance from EBS volumes, we recommend that you use an EBS-optimized instance. EBS-optimized instances deliver dedicated throughput between Amazon EC2 and Amazon EBS, with instance. EBS-optimized instances deliver dedicated bandwidth between Amazon EC2 and Amazon EBS, with specifications depending on the instance type. For more information, see Amazon EBS–Optimized Instances (p. 876).

To create an EBS-optimized instance, choose Launch as an EBS-Optimized Instance when launching the instance using the Amazon EC2 console, or specify --ebs-optimized when using the command line. Be sure that you launch a current-generation instance that supports this option. For more information, see Amazon EBS–Optimized Instances (p. 876).

Setting up Provisioned IOPS SSD (io1) volumes

To create an io1 volume, choose Provisioned IOPS SSD when creating the volume using the Amazon EC2 console, or, at the command line, specify --type io1 --iops \( n \) where \( n \) is an integer between 100 and 64,000. For more detailed EBS-volume specifications, see Amazon EBS Volume Types (p. 786). For information about creating an EBS volume, see Creating an Amazon EBS Volume (p. 800). For information about attaching a volume to an instance, see Attaching an Amazon EBS Volume to an Instance (p. 803).

For the example tests, we recommend that you create a RAID array with 6 volumes, which offers a high level of performance. Because you are charged by gigabytes provisioned (and the number of provisioned IOPS for io1 volumes), not the number of volumes, there is no additional cost for creating multiple, smaller volumes and using them to create a stripe set. If you're using Oracle Orion to benchmark your volumes, it can simulate striping the same way that Oracle ASM does, so we recommend that you let Orion do the striping. If you are using a different benchmarking tool, you need to stripe the volumes yourself.

Setting up Throughput Optimized HDD (st1) or Cold HDD (sc1) volumes

To create an st1 volume, choose Throughput Optimized HDD when creating the volume using the Amazon EC2 console, or specify --type st1 when using the command line. To create an sc1 volume, choose Cold HDD when creating the volume using the Amazon EC2 console, or specify --type sc1 when using the command line. For information about creating EBS volumes, see Creating an Amazon EBS Volume (p. 800). For information about attaching these volumes to your instance, see Attaching an Amazon EBS Volume to an Instance (p. 803).

Install Benchmark Tools

The following table lists some of the possible tools you can use to benchmark the performance of EBS volumes.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fio</td>
<td>For benchmarking I/O performance. (Note that fio has a dependency on libaio-devel.)</td>
</tr>
<tr>
<td></td>
<td>To install fio on Amazon Linux, run the following command:</td>
</tr>
<tr>
<td></td>
<td>[ec2-user ~]# sudo yum install -y fio</td>
</tr>
<tr>
<td></td>
<td>To install fio on Ubuntu, run the following command:</td>
</tr>
<tr>
<td></td>
<td>sudo apt-get install -y fio</td>
</tr>
</tbody>
</table>
## Oracle Orion Calibration Tool

For calibrating the I/O performance of storage systems to be used with Oracle databases.

These benchmarking tools support a wide variety of test parameters. You should use commands that approximate the workloads your volumes will support. These commands provided below are intended as examples to help you get started.

### Choosing the Volume Queue Length

Choosing the best volume queue length based on your workload and volume type.

#### Queue Length on SSD-backed Volumes

To determine the optimal queue length for your workload on SSD-backed volumes, we recommend that you target a queue length of 1 for every 1000 IOPS available (baseline for gp2 volumes and the provisioned amount for io1 volumes). Then you can monitor your application performance and tune that value based on your application requirements.

Increasing the queue length is beneficial until you achieve the provisioned IOPS, throughput or optimal system queue length value, which is currently set to 32. For example, a volume with 3,000 provisioned IOPS should target a queue length of 3. You should experiment with tuning these values up or down to see what performs best for your application.

#### Queue Length on HDD-backed Volumes

To determine the optimal queue length for your workload on HDD-backed volumes, we recommend that you target a queue length of at least 4 while performing 1MiB sequential I/Os. Then you can monitor your application performance and tune that value based on your application requirements. For example, a 2 TiB st1 volume with burst throughput of 500 MiB/s and IOPS of 500 should target a queue length of 4, 8, or 16 while performing 1,024 KiB, 512 KiB, or 256 KiB sequential I/Os respectively. You should experiment with tuning these values up or down to see what performs best for your application.

### Disable C-States

Before you run benchmarking, you should disable processor C-states. Temporarily idle cores in a supported CPU can enter a C-state to save power. When the core is called on to resume processing, a certain amount of time passes until the core is again fully operational. This latency can interfere with processor benchmarking routines. For more information about C-states and which EC2 instance types support them, see Processor State Control for Your EC2 Instance.

#### Disabling C-States on a Windows System

You can disable C-states on Windows as follows:

1. In PowerShell, get the current active power scheme.

   ```
   C:\> $current_scheme = powercfg /getactivescheme
   ```

2. Get the power scheme GUID.

   ```
   C:\> (Get-WmiObject -class Win32_PowerPlan -Namespace "root\cimv2\power" -Filter "ElementName='High performance'").InstanceID
   ```

3. Get the power setting GUID.
4. Get the power setting subgroup GUID.

```
C:\> (Get-WmiObject -class Win32_PowerSettingSubgroup -Namespace "root\cimv2\power" -Filter "ElementName='Processor power management'").InstanceID
```

5. Disable C-states by setting the value of the index to 1. A value of 0 indicates that C-states are disabled.

```
C:\> powercfg /setacvalueindex <power_scheme_guid> <power_setting_subgroup_guid> <power_setting_guid> 1
```

6. Set active scheme to ensure the settings are saved.

```
C:\> powercfg /setactive <power_scheme_guid>
```

**Perform Benchmarking**

The following procedures describe benchmarking commands for various EBS volume types.

Run the following commands on an EBS-optimized instance with attached EBS volumes. If the EBS volumes were restored from snapshots, be sure to initialize them before benchmarking. For more information, see Initializing Amazon EBS Volumes (p. 891).

When you are finished testing your volumes, see the following topics for help cleaning up: Deleting an Amazon EBS Volume (p. 818) and Terminate Your Instance (p. 388).

**Benchmarking io1 Volumes**

Run `fio` on the stripe set that you created.

The following command performs 16 KB random write operations.

```
[ec2-user ~]# sudo fio --directory=/mnt/p_iops_vol0 --name fio_test_file --direct=1 --rw=randwrite --bs=16k --size=1G --numjobs=16 --time_based --runtime=180 --group_reporting --norandommap
```

The following command performs 16 KB random read operations.

```
[ec2-user ~]# sudo fio --directory=/mnt/p_iops_vol0 --name fio_test_file --direct=1 --rw=randread --bs=16k --size=1G --numjobs=16 --time_based --runtime=180 --group_reporting --norandommap
```

For more information about interpreting the results, see this tutorial: Inspecting disk IO performance with fio.

**Benchmarking st1 and sc1 Volumes**

Run `fio` on your st1 or sc1 volume.

The following command performs 1 MiB sequential read operations against an attached st1 block device (e.g., `/dev/xvdf`):
The following command performs 1 MiB sequential write operations against an attached st1 block device:

```
[ec2-user ~]$ sudo fio --filename=/dev/<device> --direct=1 --rw=write --randrepeat=0 --ioengine=libaio --bs=1024k --iodepth=8 --time_based=1 --runtime=180 -- name=fio_direct_write_test
```

Some workloads perform a mix of sequential reads and sequential writes to different parts of the block device. To benchmark such a workload, we recommend that you use separate, simultaneous fio jobs for reads and writes, and use the fio offset_increment option to target different block device locations for each job.

Running this workload is a bit more complicated than a sequential-write or sequential-read workload. Use a text editor to create a fio job file, called fio_rw_mix.cfg in this example, that contains the following:

```
[globals]
clocksource=clock_gettime
randrepeat=0
runtime=180
offset_increment=100g

[sequential-write]
bs=1M
ioengine=libaio
direct=1
iodepth=8
filename=/dev/<device>
do_verify=0
rw=write
rwmixread=0
rwmixwrite=100

[sequential-read]
bs=1M
ioengine=libaio
direct=1
iodepth=8
filename=/dev/<device>
do_verify=0
rw=read
rwmixread=100
rwmixwrite=0
```

Then run the following command:

```
[ec2-user ~]$ sudo fio fio_rw_mix.cfg
```

For more information about interpreting the results, see this tutorial: Inspecting disk I/O performance with fio.

Multiple fio jobs for direct I/O, even though using sequential read or write operations, can result in lower than expected throughput for st1 and sc1 volumes. We recommend that you use one direct I/O job and use the iodepth parameter to control the number of concurrent I/O operations.
Amazon CloudWatch Metrics for Amazon EBS

CloudWatch metrics are statistical data that you can use to view, analyze, and set alarms on the operational behavior of your volumes.

The following table describes the types of monitoring data available for your Amazon EBS volumes.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Data is available automatically in 5-minute periods at no charge. This includes data for the root device volumes for EBS-backed instances.</td>
</tr>
<tr>
<td>Detailed</td>
<td>Provisioned IOPS SSD (io1) volumes automatically send one-minute metrics to CloudWatch.</td>
</tr>
</tbody>
</table>

When you get data from CloudWatch, you can include a `Period` request parameter to specify the granularity of the returned data. This is different than the period that we use when we collect the data (5-minute periods). We recommend that you specify a period in your request that is equal to or larger than the collection period to ensure that the returned data is valid.

You can get the data using either the CloudWatch API or the Amazon EC2 console. The console takes the raw data from the CloudWatch API and displays a series of graphs based on the data. Depending on your needs, you might prefer to use either the data from the API or the graphs in the console.

Amazon EBS Metrics

Amazon Elastic Block Store (Amazon EBS) sends data points to CloudWatch for several metrics. Amazon EBS General Purpose SSD (gp2), Throughput Optimized HDD (st1), Cold HDD (sc1), and Magnetic (standard) volumes automatically send five-minute metrics to CloudWatch. Provisioned IOPS SSD (io1) volumes automatically send one-minute metrics to CloudWatch. Data is only reported to CloudWatch when the volume is attached to an instance.

Some of these metrics have differences on Nitro-based instances. For a list of instance types based on the Nitro system, see Nitro-based Instances (p. 122).

The AWS/EBS namespace includes the following metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VolumeReadBytes</td>
<td>Provides information on the read operations in a specified period of time. The <code>Sum</code> statistic reports the total number of bytes transferred during the period. The <code>Average</code> statistic reports the average size of each read operation during the period, except on volumes attached to a Nitro-based instance, where the average represents the average over the specified period. The <code>SampleCount</code> statistic reports the total number of read operations during the period, except on volumes attached to a Nitro-based instance, where the sample count represents the number of data points used in the statistical calculation. For Xen instances, data is reported only when there is read activity on the volume. The <code>Minimum</code> and <code>Maximum</code> statistics on this metric are supported only by volumes attached to Nitro-based instances. Units: Bytes</td>
</tr>
</tbody>
</table>

903
<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VolumeWriteBytes</td>
<td>Provides information on the write operations in a specified period of time. The <code>Sum</code> statistic reports the total number of bytes transferred during the period. The <code>Average</code> statistic reports the average size of each write operation during the period, except on volumes attached to a Nitro-based instance, where the average represents the average over the specified period. The <code>SampleCount</code> statistic reports the total number of write operations during the period, except on volumes attached to a Nitro-based instance, where the sample count represents the number of data points used in the statistical calculation. For Xen instances, data is reported only when there is write activity on the volume. The <code>Minimum</code> and <code>Maximum</code> statistics on this metric are supported only by volumes attached to Nitro-based instances. Units: Bytes</td>
</tr>
<tr>
<td>VolumeReadOps</td>
<td>The total number of read operations in a specified period of time. To calculate the average read operations per second (read IOPS) for the period, divide the total read operations in the period by the number of seconds in that period. The <code>Minimum</code> and <code>Maximum</code> statistics on this metric are supported only by volumes attached to Nitro-based instances. Units: Count</td>
</tr>
<tr>
<td>VolumeWriteOps</td>
<td>The total number of write operations in a specified period of time. To calculate the average write operations per second (write IOPS) for the period, divide the total write operations in the period by the number of seconds in that period. The <code>Minimum</code> and <code>Maximum</code> statistics on this metric are supported only by volumes attached to Nitro-based instances. Units: Count</td>
</tr>
<tr>
<td>VolumeTotalReadTime</td>
<td>The total number of seconds spent by all read operations that completed in a specified period of time. If multiple requests are submitted at the same time, this total could be greater than the length of the period. For example, for a period of 5 minutes (300 seconds): if 700 operations completed during that period, and each operation took 1 second, the value would be 700 seconds. For Xen instances, data is reported only when there is read activity on the volume. The <code>Average</code> statistic on this metric is not relevant for volumes attached to Nitro-based instances. The <code>Minimum</code> and <code>Maximum</code> statistics on this metric are supported only by volumes attached to Nitro-based instances. Units: Seconds</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>VolumeTotalWriteTime</td>
<td>The total number of seconds spent by all write operations that completed in a specified period of time. If multiple requests are submitted at the same time, this total could be greater than the length of the period. For example, for a period of 5 minutes (300 seconds): if 700 operations completed during that period, and each operation took 1 second, the value would be 700 seconds. For Xen instances, data is reported only when there is write activity on the volume. The <code>Average</code> statistic on this metric is not relevant for volumes attached to Nitro-based instances. The <code>Minimum</code> and <code>Maximum</code> statistics on this metric are supported only by volumes attached to Nitro-based instances. Units: Seconds</td>
</tr>
<tr>
<td>VolumeIdleTime</td>
<td>The total number of seconds in a specified period of time when no read or write operations were submitted. The <code>Average</code> statistic on this metric is not relevant for volumes attached to Nitro-based instances. The <code>Minimum</code> and <code>Maximum</code> statistics on this metric are supported only by volumes attached to Nitro-based instances. Units: Seconds</td>
</tr>
</tbody>
</table>
### EBS CloudWatch Metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VolumeConsumedReadWriteOps</td>
<td>Used with Provisioned IOPS SSD volumes only. The total amount of read and write operations (normalized to 256K capacity units) consumed in a specified period of time. I/O operations that are smaller than 256K each count as 1 consumed IOPS. I/O operations that are larger than 256K are counted in 256K capacity units. For example, a 1024K I/O would count as 4 consumed IOPS. Units: Count</td>
</tr>
<tr>
<td>BurstBalance</td>
<td>Used with General Purpose SSD (gp2), Throughput Optimized HDD (st1), and Cold HDD (sc1) volumes only. Provides information about the percentage of I/O credits (for gp2) or throughput credits (for st1 and sc1) remaining in the burst bucket. Data is reported to CloudWatch only when the volume is active. If the volume is not attached, no data is reported. The Sum statistic on this metric is not relevant for volumes attached to Nitro-based instances. For a volume 1 TiB or larger, baseline performance is higher than maximum burst performance, so I/O credits are never spent. If the volume is attached to a Nitro-based instance, the reported burst balance is 0%. For a non-Nitro-based instance, the reported burst balance is 100%. Units: Percent</td>
</tr>
</tbody>
</table>

### Dimensions for Amazon EBS Metrics

The supported dimension is the volume ID (VolumeId). All available statistics are filtered by volume ID.

### Graphs in the Amazon EC2 Console

After you create a volume, you can view the volume's monitoring graphs in the Amazon EC2 console. Select a volume on the Volumes page in the console and choose Monitoring. The following table lists the graphs that are displayed. The column on the right describes how the raw data metrics from the CloudWatch API are used to produce each graph. The period for all the graphs is 5 minutes.

<table>
<thead>
<tr>
<th>Graph</th>
<th>Description using raw metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Bandwidth (KiB/s)</td>
<td>Sum(VolumeReadBytes) / Period / 1024</td>
</tr>
<tr>
<td>Write Bandwidth (KiB/s)</td>
<td>Sum(VolumeWriteBytes) / Period / 1024</td>
</tr>
<tr>
<td>Read Throughput (IOPS)</td>
<td>Sum(VolumeReadOps) / Period</td>
</tr>
<tr>
<td>Write Throughput (IOPS)</td>
<td>Sum(VolumeWriteOps) / Period</td>
</tr>
<tr>
<td>Avg Queue Length (Operations)</td>
<td>Avg(VolumeQueueLength)</td>
</tr>
<tr>
<td>% Time Spent Idle</td>
<td>Sum(VolumeIdleTime) / Period × 100</td>
</tr>
<tr>
<td>Avg Read Size (KiB/Operation)</td>
<td>Avg(VolumeReadBytes) / 1024</td>
</tr>
</tbody>
</table>
### Description using raw metrics

For Nitro-based instances, the following formula derives Average Read Size using CloudWatch Metric Math:

\[
\frac{\text{Sum(VolumeReadBytes)}}{\text{Sum(VolumeReadOps)}} / 1024
\]

The VolumeReadBytes and VolumeReadOps metrics are available in the EBS CloudWatch console.

### Avg Write Size (KiB/Operation)

\[
\frac{\text{Avg(VolumeWriteBytes)}}{1024}
\]

For Nitro-based instances, the following formula derives Average Write Size using CloudWatch Metric Math:

\[
\frac{\text{Sum(VolumeWriteBytes)}}{\text{Sum(VolumeWriteOps)}} / 1024
\]

The VolumeWriteBytes and VolumeWriteOps metrics are available in the EBS CloudWatch console.

### Avg Read Latency (ms/Operation)

\[
\frac{\text{Avg(VolumeTotalReadTime)}}{1000}
\]

For Nitro-based instances, the following formula derives Average Read Latency using CloudWatch Metric Math:

\[
\frac{\text{Sum(VolumeTotalReadTime)}}{\text{Sum(VolumeReadOps)}} × 1000
\]

The VolumeTotalReadTime and VolumeReadOps metrics are available in the EBS CloudWatch console.

### Avg Write Latency (ms/Operation)

\[
\frac{\text{Avg(VolumeTotalWriteTime)}}{1000}
\]

For Nitro-based instances, the following formula derives Average Write Latency using CloudWatch Metric Math:

\[
\frac{\text{Sum(VolumeTotalWriteTime)}}{\text{Sum(VolumeWriteOps)}} × 1000
\]

The VolumeTotalWriteTime and VolumeWriteOps metrics are available in the EBS CloudWatch console.

For the average latency graphs and average size graphs, the average is calculated over the total number of operations (read or write, whichever is applicable to the graph) that completed during the period.

### Amazon CloudWatch Events for Amazon EBS

Amazon EBS emits notifications based on Amazon CloudWatch Events for a variety of volume, snapshot, and encryption status changes. With CloudWatch Events, you can establish rules that trigger programmatic actions in response to a change in volume, snapshot, or encryption key state. For example, when a snapshot is created, you can trigger an AWS Lambda function to share the completed snapshot with another account or copy it to another region for disaster-recovery purposes.

Events in CloudWatch are represented as JSON objects. The fields that are unique to the event are contained in the "detail" section of the JSON object. The "event" field contains the event name. The "result" field contains the completed status of the action that triggered the event. For more information, see Event Patterns in CloudWatch Events in the Amazon CloudWatch Events User Guide.
For more information, see Using Events in the Amazon CloudWatch User Guide.

Contents
- EBS Volume Events (p. 908)
- EBS Snapshot Events (p. 911)
- EBS Volume Modification Events (p. 914)
- Using Amazon Lambda To Handle CloudWatch Events (p. 915)

EBS Volume Events

Amazon EBS sends events to CloudWatch Events when the following volume events occur.

Events
- Create Volume (createVolume) (p. 908)
- Delete Volume (deleteVolume) (p. 909)
- Volume Attach or Reattach (attachVolume, reattachVolume) (p. 910)

Create Volume (createVolume)

The createVolume event is sent to your AWS account when an action to create a volume completes. However it is not saved, logged, or archived. This event can have a result of either available or failed. Creation will fail if an invalid KMS key was provided, as shown in the examples below.

Event Data

The listing below is an example of a JSON object emitted by EBS for a successful createVolume event.

```json
{
  "version": "0",
  "id": "01234567-0123-0123-0123-012345678901",
  "detail-type": "EBS Volume Notification",
  "source": "aws.ec2",
  "account": "012345678901",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-east-1",
  "resources": [
    "arn:aws:ec2:us-east-1:012345678901:volume/vol-01234567"
  ],
  "detail": {
    "result": "available",
    "cause": "",
    "event": "createVolume",
    "request-id": "01234567-0123-0123-0123-0123456789ab"
  }
}
```

The listing below is an example of a JSON object emitted by EBS after a failed createVolume event. The cause for the failure was a disabled KMS key.

```json
{
  "version": "0",
  "id": "01234567-0123-0123-0123-0123456789ab",
  "detail-type": "EBS Volume Notification",
  "source": "aws.ec2",
  "account": "012345678901",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "sa-east-1",
  "resources": [
  ],
  "detail": {
    "result": "available",
    "cause": "",
    "event": "createVolume",
    "request-id": "01234567-0123-0123-0123-0123456789ab"
  }
}
```
The following is an example of a JSON object that is emitted by EBS after a failed createVolume event. The cause for the failure was a KMS key pending import.

```
{
  "version": "0",
  "id": "01234567-0123-0123-0123-0123456789ab",
  "detail-type": "EBS Volume Notification",
  "source": "aws.ec2",
  "account": "012345678901",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-east-1",
  "resources": [
  ],
  "detail": {
    "event": "createVolume",
    "result": "failed",
    "cause": "arn:aws:kms:sa-east-1:0123456789ab:key/01234567-0123-0123-0123456789ab is pending import.",
    "request-id": "01234567-0123-0123-0123-0123456789ab",
  }
}
```

Delete Volume (deleteVolume)

The deleteVolume event is sent to your AWS account when an action to delete a volume completes. However it is not saved, logged, or archived. This event has the result deleted. If the deletion does not complete, the event is never sent.

Event Data

The listing below is an example of a JSON object emitted by EBS for a successful deleteVolume event.

```
{
  "version": "0",
  "id": "01234567-0123-0123-0123-012345678901",
  "detail-type": "EBS Volume Notification",
  "source": "aws.ec2",
  "account": "012345678901",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-east-1",
  "resources": [
    "arn:aws:ec2:us-east-1:012345678901:volume/vol-0123456789ab"
  ],
  "detail": {
    "result": "deleted",
    "cause": "",
    "event": "deleteVolume",
    "request-id": "01234567-0123-0123-0123-0123456789ab"
  }
}
```
Volume Attach or Reattach (attachVolume, reattachVolume)

The attachVolume or reattachVolume event is sent to your AWS account if a volume fails to attach or reattach to an instance. However it is not saved, logged, or archived. If you use a KMS key to encrypt an EBS volume and the key becomes invalid, EBS will emit an event if that key is later used to attach or reattach to an instance, as shown in the examples below.

Event Data

The listing below is an example of a JSON object emitted by EBS after a failed attachVolume event. The cause for the failure was a KMS key pending deletion.

Note

AWS may attempt to reattach to a volume following routine server maintenance.

```json
{
   "version": "0",
   "id": "01234567-0123-0123-0123-0123456789ab",
   "detail-type": "EBS Volume Notification",
   "source": "aws.ec2",
   "account": "012345678901",
   "time": "yyyyMMddThhmmssZ",
   "region": "us-east-1",
   "resources": [
      "arn:aws:kms:us-east-1:0123456789ab:key/01234567-0123-0123-0123456789ab"
   ],
   "detail": {
      "event": "attachVolume",
      "result": "failed",
      "cause": "arn:aws:kms:us-east-1:0123456789ab:key/01234567-0123-0123-0123456789ab is pending deletion."
   }
}
```

The listing below is an example of a JSON object emitted by EBS after a failed reattachVolume event. The cause for the failure was a KMS key pending deletion.

```json
{
   "version": "0",
   "id": "01234567-0123-0123-0123-0123456789ab",
   "detail-type": "EBS Volume Notification",
   "source": "aws.ec2",
   "account": "012345678901",
   "time": "yyyyMMddThhmmssZ",
   "region": "us-east-1",
   "resources": [
      "arn:aws:kms:us-east-1:0123456789ab:key/01234567-0123-0123-0123456789ab"
   ],
   "detail": {
      "event": "reattachVolume",
      "result": "failed",
      "cause": "arn:aws:kms:us-east-1:0123456789ab:key/01234567-0123-0123-0123456789ab is pending deletion."
   }
}
```
EBS Snapshot Events

Amazon EBS sends events to CloudWatch Events when the following volume events occur.

Events
- Create Snapshot (createSnapshot) (p. 911)
- Create Snapshots (createSnapshots) (p. 911)
- Copy Snapshot (copySnapshot) (p. 913)
- Share Snapshot (shareSnapshot) (p. 914)

Create Snapshot (createSnapshot)

The createSnapshot event is sent to your AWS account when an action to create a snapshot completes. However it is not saved, logged, or archived. This event can have a result of either succeeded or failed.

Event Data

The listing below is an example of a JSON object emitted by EBS for a successful createSnapshot event. In the detail section, the source field contains the ARN of the source volume. The StartTime and EndTime fields indicate when creation of the snapshot started and completed.

```
{
"version": "0",
"id": "01234567-0123-0123-0123-012345678901",
"detail-type": "EBS Snapshot Notification",
"source": "aws.ec2",
"account": "012345678901",
"time": "yyyy-mm-ddThh:mm:ssZ",
"region": "us-east-1",
"resources": [
  "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
],
"detail": {
  "event": "createSnapshot",
  "result": "succeeded",
  "cause": "",
  "request-id": "",
  "snapshot_id": "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
  "source": "arn:aws:ec2:us-west-2::volume/vol-01234567",
  "StartTime": "yyyy-mm-ddThh:mm:ssZ",
  "EndTime": "yyyy-mm-ddThh:mm:ssZ"
}
```

Create Snapshots (createSnapshots)

The createSnapshots event is sent to your AWS account when an action to create a multi-volume snapshot completes. This event can have a result of either succeeded or failed.

Event Data

The listing below is an example of a JSON object emitted by EBS for a successful createSnapshots event. In the detail section, the source field contains the ARNs of the source volumes of the multi-volume snapshot set. The StartTime and EndTime fields indicate when creation of the snapshot started and completed.

```
{
```

911
The listing below is an example of a JSON object emitted by EBS after a failed createSnapshots event. The cause for the failure was one or more snapshots failed to complete. The values of snapshot_id are the ARNs of the failed snapshots. StartTime and EndTime represent when the create-snapshots action started and ended.

```json
{
  "version": "0",
  "id": "01234567-0123-0123-0123-012345678901",
  "detail-type": "EBS Multi-Volume Snapshots Completion Status",
  "source": "aws.ec2",
  "account": "012345678901",
  "time": "yyyyMMddThhmmssZ",
  "region": "us-east-1",
  "resources": [
    "arn:aws:ec2::us-east-1:snapshot/snap-01234567",
    "arn:aws:ec2::us-east-1:snapshot/snap-012345678"
  ],
  "detail": {
    "event": "createSnapshots",
    "result": "succeeded",
    "cause": "",
    "request-id": "",
    "startTime": "yyyyMMddThhmmssZ",
    "endTime": "yyyyMMddThhmmssZ",
    "snapshots": [
      {
        "snapshot_id": "arn:aws:ec2::us-east-1:snapshot/snap-01234567",
        "source": "arn:aws:ec2::us-east-1:volume/vol-01234567",
        "status": "completed"
      },
      {
        "snapshot_id": "arn:aws:ec2::us-east-1:snapshot/snap-012345678",
        "source": "arn:aws:ec2::us-east-1:volume/vol-012345678",
        "status": "completed"
      }
    ]
  }
}
```
Copy Snapshot (copySnapshot)

The copySnapshot event is sent to your AWS account when an action to copy a snapshot completes. However, it is not saved, logged, or archived. This event can have a result of either succeeded or failed.

Event Data

The listing below is an example of a JSON object emitted by EBS after a successful copySnapshot event. The value of snapshot_id is the ARN of the newly created snapshot. In the detail section, the value of source is the ARN of the source snapshot. StartTime and EndTime represent when the copy-snapshot action started and ended.

```
{
    "version": "0",
    "id": "01234567-0123-0123-0123-012345678901",
    "detail-type": "EBS Snapshot Notification",
    "source": "aws.ec2",
    "account": "123456789012",
    "time": "yyyymmddThhmmssZ",
    "region": "us-east-1",
    "resources": [
        "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
        "arn:aws:ec2:us-east-1::snapshot/snap-012345678901",
        "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
        "arn:aws:ec2:us-east-1::snapshot/snap-012345678901"
    ],
    "detail": {
        "event": "copySnapshot",
        "result": "succeeded",
        "cause": "",
        "request-id": "",
        "snapshot_id": "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
        "source": "arn:aws:ec2:eu-west-1::snapshot/snap-76543210",
        "StartTime": "yyyymmddThhmmssZ",
        "EndTime": "yyyymmddThhmmssZ",
        "Incremental": "True"
    }
}
```

The listing below is an example of a JSON object emitted by EBS after a failed copySnapshot event. The cause for the failure was an invalid source snapshot ID. The value of snapshot_id is the ARN of the failed snapshot. In the detail section, the value of source is the ARN of the source snapshot. StartTime and EndTime represent when the copy-snapshot action started and ended.

```
{
    "version": "0",
    "id": "01234567-0123-0123-0123-012345678901",
    "detail-type": "EBS Snapshot Notification",
    "source": "aws.ec2",
    "account": "123456789012",
    "time": "yyyymmddThhmmssZ",
    "region": "us-east-1",
    "resources": [
        "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
        "arn:aws:ec2:us-east-1::snapshot/snap-012345678901"
    ],
    "detail": {
        "event": "copySnapshot",
        "result": "failed",
        "cause": "Invalid snapshot ID",
        "request-id": "",
        "snapshot_id": "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
        "source": "arn:aws:ec2:us-east-1::snapshot/snap-012345678901",
        "StartTime": "yyyymmddThhmmssZ",
        "EndTime": "yyyymmddThhmmssZ"
    }
}
```
Share Snapshot (shareSnapshot)

The shareSnapshot event is sent to your AWS account when another account shares a snapshot with it. However it is not saved, logged, or archived. The result is always succeeded.

**Event Data**

The following is an example of a JSON object emitted by EBS after a completed shareSnapshot event. In the `detail` section, the value of `source` is the AWS account number of the user that shared the snapshot with you. `StartTime` and `EndTime` represent when the share-snapshot action started and ended. The shareSnapshot event is emitted only when a private snapshot is shared with another user. Sharing a public snapshot does not trigger the event.

```
{
  "version": "0",
  "id": "01234567-0123-0123-0123-012345678901",
  "detail-type": "EBS Snapshot Notification",
  "source": "aws.ec2",
  "account": "012345678901",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-east-1",
  "resources": [
    "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
  ],
  "detail": {
    "event": "shareSnapshot",
    "result": "succeeded",
    "cause": "",
    "request-id": "",
    "snapshot_id": "arn:aws:ec2:us-west-2::snapshot/snap-01234567",
    "source": "012345678901",
    "StartTime": "yyyy-mm-ddThh:mm:ssZ",
    "EndTime": "yyyy-mm-ddThh:mm:ssZ"
  }
}
```

EBS Volume Modification Events

Amazon EBS sends modifyVolume events to CloudWatch Events when a volume is modified. However it is not saved, logged, or archived.

```
{
  "version": "0",
  "id": "01234567-0123-0123-0123-012345678901",
  "detail-type": "EBS Volume Notification",
  "source": "aws.ec2",
  "account": "012345678901",
  "time": "yyyy-mm-ddThh:mm:ssZ",
}
```
Using Amazon Lambda To Handle CloudWatch Events

You can use Amazon EBS and CloudWatch Events to automate your data-backup workflow. This requires you to create an IAM policy, a AWS Lambda function to handle the event, and an Amazon CloudWatch Events rule that matches incoming events and routes them to the Lambda function.

The following procedure uses the createSnapshot event to automatically copy a completed snapshot to another region for disaster recovery.

To copy a completed snapshot to another region

1. Create an IAM policy, such as the one shown in the following example, to provide permissions to execute a CopySnapshot action and write to the CloudWatch Events log. Assign the policy to the IAM user that will handle the CloudWatch event.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": "arn:aws:logs:*:*:*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CopySnapshot"
            ],
            "Resource": "*"
        }
    ]
}
```

2. Define a function in Lambda that will be available from the CloudWatch console. The sample Lambda function below, written in Node.js, is invoked by CloudWatch when a matching createSnapshot event is emitted by Amazon EBS (signifying that a snapshot was completed). When invoked, the function copies the snapshot from us-east-2 to us-east-1.

```javascript
// Sample Lambda function to copy an EBS snapshot to a different region

var AWS = require('aws-sdk');
var ec2 = new AWS.EC2();

// define variables
var destinationRegion = 'us-east-1';

// create a new instance
var instance = ec2.instances.createSnapshot()
    .promise()
    .then(function(snapshot) {
        console.log('Snapshot created: ' + snapshot.createSnapshotResult.snapshotId);
    })
    .catch(function(err) {
        console.log('Error creating snapshot: ' + err);
    });
```

Using Amazon Elastic Compute Cloud
User Guide for Windows Instances
EBS CloudWatch Events
```javascript
var sourceRegion = 'us-east-2';

console.log('Loading function');

//main function
exports.handler = (event, context, callback) => {
    // Get the EBS snapshot ID from the CloudWatch event details
    var snapshotArn = event.detail.snapshot_id.split('/');
    const snapshotId = snapshotArn[1];
    const description = `Snapshot copy from ${snapshotId} in ${sourceRegion}.`; 
    console.log(`snapshotId: ${snapshotId}`);

    // Load EC2 class and update the configuration to use destination region to initiate the snapshot.
    AWS.config.update({region: destinationRegion});

    var ec2 = new AWS.EC2();

    // Prepare variables for ec2.modifySnapshotAttribute call
    const copySnapshotParams = {
        Description: description,
        DestinationRegion: destinationRegion,
        SourceRegion: sourceRegion,
        SourceSnapshotId: snapshotId
    };

    // Execute the copy snapshot and log any errors
    ec2.copySnapshot(copySnapshotParams, (err, data) => {
        if (err) {
            const errorMessage = `Error copying snapshot ${snapshotId} to region ${destinationRegion}.`;
            console.log(errorMessage);
            console.log(err);
            callback(errorMessage);
        } else {
            const successMessage = `Successfully started copy of snapshot ${snapshotId} to region ${destinationRegion}.`;
            console.log(successMessage);
            console.log(data);
            callback(null, successMessage);
        }
    });
};
```

To ensure that your Lambda function is available from the CloudWatch console, create it in the region where the CloudWatch event will occur. For more information, see the AWS Lambda Developer Guide.

4. Choose Events, Create rule, Select event source, and Amazon EBS Snapshots.
5. For Specific Event(s), choose createSnapshot and for Specific Result(s), choose succeeded.
6. For Rule target, find and choose the sample function that you previously created.
7. Choose Target, Add Target.
8. For Lambda function, select the Lambda function that you previously created and choose Configure details.
9. On the Configure rule details page, type values for Name and Description. Select the State check box to activate the function (setting it to Enabled).
10. Choose Create rule.

Your rule should now appear on the Rules tab. In the example shown, the event that you configured should be emitted by EBS the next time you copy a snapshot.
Amazon EC2 Instance Store

An instance store provides temporary block-level storage for your instance. This storage is located on disks that are physically attached to the host computer. Instance store is ideal for temporary storage of information that changes frequently, such as buffers, caches, scratch data, and other temporary content, or for data that is replicated across a fleet of instances, such as a load-balanced pool of web servers.

An instance store consists of one or more instance store volumes exposed as block devices. The size of an instance store as well as the number of devices available varies by instance type.

The virtual devices for instance store volumes are ephemeral[0–23]. Instance types that support one instance store volume have ephemeral0. Instance types that support two instance store volumes have ephemeral0 and ephemeral1, and so on.

Instance Store Lifetime

You can specify instance store volumes for an instance only when you launch it. You can't detach an instance store volume from one instance and attach it to a different instance.

The data in an instance store persists only during the lifetime of its associated instance. If an instance reboots (intentionally or unintentionally), data in the instance store persists. However, data in the instance store is lost under any of the following circumstances:

• The underlying disk drive fails
• The instance stops
• The instance terminates

Therefore, do not rely on instance store for valuable, long-term data. Instead, use more durable data storage, such as Amazon S3, Amazon EBS, or Amazon EFS.
When you stop or terminate an instance, every block of storage in the instance store is reset. Therefore, your data cannot be accessed through the instance store of another instance.

If you create an AMI from an instance, the data on its instance store volumes isn't preserved and isn't present on the instance store volumes of the instances that you launch from the AMI.

If you change the instance type, an instance store will not be attached to the new instance type. For more information, see Changing the Instance Type (p. 188).

**Instance Store Volumes**

The instance type determines the size of the instance store available and the type of hardware used for the instance store volumes. Instance store volumes are included as part of the instance's usage cost. You must specify the instance store volumes that you'd like to use when you launch the instance (except for NVMe instance store volumes, which are available by default). Then format and mount the instance store volumes before using them. You can't make an instance store volume available after you launch the instance. For more information, see Add Instance Store Volumes to Your EC2 Instance (p. 922).

Some instance types use NVMe or SATA-based solid state drives (SSD) to deliver high random I/O performance. This is a good option when you need storage with very low latency, but you don't need the data to persist when the instance terminates or you can take advantage of fault-tolerant architectures. For more information, see SSD Instance Store Volumes (p. 924).

The following table provides the quantity, size, type, and performance optimizations of instance store volumes available on each supported instance type. For a complete list of instance types, including EBS-only types, see Amazon EC2 Instance Types.

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Instance Store Volumes</th>
<th>Type</th>
<th>Needs Initialization*</th>
<th>TRIM Support**</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1.medium</td>
<td>1 x 350 GB</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c1.xlarge</td>
<td>4 x 420 GB (1.6 TB)</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c3.large</td>
<td>2 x 16 GB (32 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c3.xlarge</td>
<td>2 x 40 GB (80 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c3.2xlarge</td>
<td>2 x 80 GB (160 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c3.4xlarge</td>
<td>2 x 160 GB (320 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c3.8xlarge</td>
<td>2 x 320 GB (640 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c5d.large</td>
<td>1 x 50 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c5d.xlarge</td>
<td>1 x 100 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c5d.2xlarge</td>
<td>1 x 200 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c5d.4xlarge</td>
<td>1 x 400 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c5d.9xlarge</td>
<td>1 x 900 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>c5d.18xlarge</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>cc2.8xlarge</td>
<td>4 x 840 GB (3.36 TB)</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>cr1.8xlarge</td>
<td>2 x 120 GB (240 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>d2.xlarge</td>
<td>3 x 2,000 GB (6 TB)</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Instance Type</td>
<td>Instance Store Volumes</td>
<td>Type</td>
<td>Needs Initialization*</td>
<td>TRIM Support**</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>------------</td>
<td>------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>d2.2xlarge</td>
<td>6 x 2,000 GB (12 TB)</td>
<td>HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d2.4xlarge</td>
<td>12 x 2,000 GB (24 TB)</td>
<td>HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d2.8xlarge</td>
<td>24 x 2,000 GB (48 TB)</td>
<td>HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f1.2xlarge</td>
<td>1 x 470 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>f1.4xlarge</td>
<td>1 x 940 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>f1.16xlarge</td>
<td>4 x 940 GB (3.76 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g2.2xlarge</td>
<td>1 x 60 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g2.8xlarge</td>
<td>2 x 120 GB (240 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g4dn.xlarge</td>
<td>1 x 125 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g4dn.2xlarge</td>
<td>1 x 225 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g4dn.4xlarge</td>
<td>1 x 225 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g4dn.8xlarge</td>
<td>1 x 900 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g4dn.12xlarge</td>
<td>1 x 900 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>g4dn.16xlarge</td>
<td>1 x 900 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>h1.2xlarge</td>
<td>1 x 2000 GB (2 TB)</td>
<td>HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h1.4xlarge</td>
<td>2 x 2000 GB (4 TB)</td>
<td>HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h1.8xlarge</td>
<td>4 x 2000 GB (8 TB)</td>
<td>HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h1.16xlarge</td>
<td>8 x 2000 GB (16 TB)</td>
<td>HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hs1.8xlarge</td>
<td>24 x 2,000 GB (48 TB)</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i2.xlarge</td>
<td>1 x 800 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i2.2xlarge</td>
<td>2 x 800 GB (1.6 TB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i2.4xlarge</td>
<td>4 x 800 GB (3.2 TB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i2.8xlarge</td>
<td>8 x 800 GB (6.4 TB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3.large</td>
<td>1 x 475 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3.xlarge</td>
<td>1 x 950 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3.2xlarge</td>
<td>1 x 1,900 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3.4xlarge</td>
<td>2 x 1,900 GB (3.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3.8xlarge</td>
<td>4 x 1,900 GB (7.6 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3.16xlarge</td>
<td>8 x 1,900 GB (15.2 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3.metal</td>
<td>8 x 1,900 GB (15.2 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Instance Type</td>
<td>Instance Store Volumes</td>
<td>Type</td>
<td>Needs Initialization*</td>
<td>TRIM Support**</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------</td>
<td>---------------</td>
<td>------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>i3en.large</td>
<td>1 x 1,250 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3en.xlarge</td>
<td>1 x 2,500 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3en.2xlarge</td>
<td>2 x 2,500 GB (5 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3en.3xlarge</td>
<td>1 x 7,500 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3en.6xlarge</td>
<td>2 x 7,500 GB (15 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3en.12xlarge</td>
<td>4 x 7,500 GB (30 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3en.24xlarge</td>
<td>8 x 7,500 GB (60 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>i3en.metal</td>
<td>8 x 7,500 GB (60 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m1.small</td>
<td>1 x 160 GB</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m1.medium</td>
<td>1 x 410 GB</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m1.large</td>
<td>2 x 420 GB (840 GB)</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m1.xlarge</td>
<td>4 x 420 GB (1.6 TB)</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m2.xlarge</td>
<td>1 x 420 GB</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m2.2xlarge</td>
<td>1 x 850 GB</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m2.4xlarge</td>
<td>2 x 840 GB (1.68 TB)</td>
<td>HDD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m3.medium</td>
<td>1 x 4 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m3.large</td>
<td>1 x 32 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m3.xlarge</td>
<td>2 x 40 GB (80 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m3.2xlarge</td>
<td>2 x 80 GB (160 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5ad.large</td>
<td>1 x 75 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5ad.xlarge</td>
<td>1 x 150 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5ad.2xlarge</td>
<td>1 x 300 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5ad.4xlarge</td>
<td>2 x 300 GB (600 GB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5ad.12xlarge</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5ad.24xlarge</td>
<td>4 x 900 GB (3.6 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.large</td>
<td>1 x 75 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.xlarge</td>
<td>1 x 150 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.2xlarge</td>
<td>1 x 300 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.4xlarge</td>
<td>2 x 300 GB (600 GB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.8xlarge</td>
<td>2 x 600 GB (1.2 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Instance Type</td>
<td>Instance Store Volumes</td>
<td>Type</td>
<td>Needs Initialization*</td>
<td>TRIM Support**</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>m5d.12xlarge</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.16xlarge</td>
<td>4 x 600 GB (2.4 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.24xlarge</td>
<td>4 x 900 GB (3.6 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>m5d.metal</td>
<td>4 x 900 GB (3.6 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>p3dn.24xlarge</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r3.large</td>
<td>1 x 32 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r3.xlarge</td>
<td>1 x 80 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r3.2xlarge</td>
<td>1 x 160 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r3.4xlarge</td>
<td>1 x 320 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r3.8xlarge</td>
<td>2 x 320 GB (640 GB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5ad.large</td>
<td>1 x 75 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5ad.xlarge</td>
<td>1 x 150 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5ad.2xlarge</td>
<td>1 x 300 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5ad.4xlarge</td>
<td>2 x 300 GB (600 GB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5ad.12xlarge</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5ad.24xlarge</td>
<td>4 x 900 GB (3.6 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.large</td>
<td>1 x 75 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.xlarge</td>
<td>1 x 150 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.2xlarge</td>
<td>1 x 300 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.4xlarge</td>
<td>2 x 300 GB (600 GB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.8xlarge</td>
<td>2 x 600 GB (1.2 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.12xlarge</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.16xlarge</td>
<td>4 x 600 GB (2.4 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.24xlarge</td>
<td>4 x 900 GB (3.6 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>r5d.metal</td>
<td>4 x 900 GB (3.6 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>x1.16xlarge</td>
<td>1 x 1,920 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>x1.32xlarge</td>
<td>2 x 1,920 GB (3.84 TB)</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>x1e.xlarge</td>
<td>1 x 120 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>x1e.2xlarge</td>
<td>1 x 240 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>x1e.4xlarge</td>
<td>1 x 480 GB</td>
<td>SSD</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Add Instance Store Volumes

### Instance Type

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Instance Store Volumes</th>
<th>Type</th>
<th>Needs Initialization*</th>
<th>TRIM Support**</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1e.8xlarge</td>
<td>1 x 960 GB</td>
<td>SSD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x1e.16xlarge</td>
<td>1 x 1,920 GB</td>
<td>SSD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x1e.32xlarge</td>
<td>2 x 1,920 GB (3.84 TB)</td>
<td>SSD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>z1d.large</td>
<td>1 x 75 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>z1d.xlarge</td>
<td>1 x 150 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>z1d.2xlarge</td>
<td>1 x 300 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>z1d.3xlarge</td>
<td>1 x 450 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>z1d.6xlarge</td>
<td>1 x 900 GB</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>z1d.12xlarge</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>z1d.metal</td>
<td>2 x 900 GB (1.8 TB)</td>
<td>NVMe SSD</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

* Volumes attached to certain instances suffer a first-write penalty unless initialized.

** For more information, see Instance Store Volume TRIM Support (p. 925).

## Add Instance Store Volumes to Your EC2 Instance

You specify the EBS volumes and instance store volumes for your instance using a block device mapping. Each entry in a block device mapping includes a device name and the volume that it maps to. The default block device mapping is specified by the AMI you use. Alternatively, you can specify a block device mapping for the instance when you launch it. All the NVMe instance store volumes supported by an instance type are automatically enumerated and assigned a device name on instance launch; including them in the block device mapping for the AMI or the instance has no effect. For more information, see Block Device Mapping (p. 930).

A block device mapping always specifies the root volume for the instance. The root volume is mounted automatically.

You can use a block device mapping to specify additional EBS volumes when you launch your instance, or you can attach additional EBS volumes after your instance is running. For more information, see Amazon EBS Volumes (p. 783).

You can specify the instance store volumes for your instance only when you launch an instance. You can't attach instance store volumes to an instance after you've launched it.

If you change the instance type, an instance store will not be attached to the new instance type. For more information, see Changing the Instance Type (p. 188).

The number and size of available instance store volumes for your instance varies by instance type. Some instance types do not support instance store volumes. For more information about the instance store volumes support by each instance type, see Instance Store Volumes (p. 918). If the instance type you choose for your instance supports instance store volumes, you must add them to the block device mapping for the instance when you launch it. After you launch the instance, you must ensure that the instance store volumes for your instance are formatted and mounted before you can use them. The root volume of an instance store-backed instance is mounted automatically.

### Contents

922
Adding Instance Store Volumes to an AMI

You can create an AMI with a block device mapping that includes instance store volumes. After you add instance store volumes to an AMI, any instance that you launch from the AMI includes these instance store volumes. When you launch an instance, you can omit volumes specified in the AMI block device mapping and add new volumes.

Important
For M3 instances, specify instance store volumes in the block device mapping of the instance, not the AMI. Amazon EC2 might ignore instance store volumes that are specified only in the block device mapping of the AMI.

To add instance store volumes to an Amazon EBS-backed AMI using the console
1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances** and select the instance.
3. Choose **Actions**, **Image**, **Create Image**.
4. In the **Create Image** dialog box, type a meaningful name and description for your image.
5. For each instance store volume to add, choose **Add New Volume**, from **Volume Type** select an instance store volume, and from **Device** select a device name. (For more information, see **Device Naming on Windows Instances** (p. 929).) The number of available instance store volumes depends on the instance type. For instances with NVMe instance store volumes, the device mapping of these volumes depends on the order in which the operating system enumerates the volumes.
6. Choose **Create Image**.

To add instance store volumes to an AMI using the command line
You can use one of the following commands. For more information about these command line interfaces, see **Accessing Amazon EC2** (p. 3).

- `create-image` or `register-image` (AWS CLI)
- `New-EC2Image` and `Register-EC2Image` (AWS Tools for Windows PowerShell)

Adding Instance Store Volumes to an Instance

When you launch an instance, the default block device mapping is provided by the specified AMI. If you need additional instance store volumes, you must add them to the instance as you launch it. You can also omit devices specified in the AMI block device mapping.

Important
For M3 instances, you might receive instance store volumes even if you do not specify them in the block device mapping for the instance.

Important
For HS1 instances, no matter how many instance store volumes you specify in the block device mapping of an AMI, the block device mapping for an instance launched from the AMI automatically includes the maximum number of supported instance store volumes. You must explicitly remove the instance store volumes that you don't want from the block device mapping for the instance before you launch it.
To update the block device mapping for an instance using the console

1. Open the Amazon EC2 console.
2. From the dashboard, choose **Launch Instance**.
3. In **Step 1: Choose an Amazon Machine Image (AMI)**, select the AMI to use and choose **Select**.
4. Follow the wizard to complete **Step 1: Choose an Amazon Machine Image (AMI)**, **Step 2: Choose an Instance Type**, and **Step 3: Configure Instance Details**.
5. In **Step 4: Add Storage**, modify the existing entries as needed. For each instance store volume to add, choose **Add New Volume**, from **Volume Type** select an instance store volume, and from **Device** select a device name. The number of available instance store volumes depends on the instance type.
6. Complete the wizard and launch the instance.

To update the block device mapping for an instance using the command line

You can use one of the following options commands with the corresponding command. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- `--block-device-mappings` with `run-instances` (AWS CLI)
- `-BlockDeviceMapping` with `New-EC2Instance` (AWS Tools for Windows PowerShell)

Making Instance Store Volumes Available on Your Instance

After you launch an instance, the instance store volumes are available to the instance, but you can’t access them until they are mounted. For Linux instances, the instance type determines which instance store volumes are mounted for you and which are available for you to mount yourself. For Windows instances, the EC2Config service mounts the instance store volumes for an instance. The block device driver for the instance assigns the actual volume name when mounting the volume, and the name assigned can be different than the name that Amazon EC2 recommends.

Many instance store volumes are pre-formatted with the ext3 file system. SSD-based instance store volumes that support TRIM instruction are not pre-formatted with any file system. However, you can format volumes with the file system of your choice after you launch your instance. For more information, see Instance Store Volume TRIM Support (p. 925). For Windows instances, the EC2Config service reformats the instance store volumes with the NTFS file system.

You can confirm that the instance store devices are available from within the instance itself using instance metadata. For more information, see Viewing the Instance Block Device Mapping for Instance Store Volumes (p. 938).

For Windows instances, you can also view the instance store volumes using Windows Disk Management. For more information, see Listing the Disks Using Windows Disk Management (p. 940).

SSD Instance Store Volumes

The following instances support instance store volumes that use solid state drives (SSD) to deliver high random I/O performance: C3, G2, I2, M3, R3, and X1. For more information about the instance store volumes support by each instance type, see Instance Store Volumes (p. 918).

Like other instance store volumes, you must map the SSD instance store volumes for your instance when you launch it. The data on an SSD instance volume persists only for the life of its associated instance. For more information, see Add Instance Store Volumes to Your EC2 Instance (p. 922).
**NVMe SSD Volumes**

The following instances offer non-volatile memory express (NVMe) SSD instance store volumes: C5d, I3, I3en, F1, M5ad, M5d, p3dn.24xlarge, R5ad, R5d, and z1d. The latest AWS Windows AMIs for the following operating systems contain the AWS NVMe drivers used to interact with SSD instance store volumes that are exposed as NVMe block devices for better performance:

- Windows Server 2019
- Windows Server 2016
- Windows Server 2012 R2
- Windows Server 2012
- Windows Server 2008 R2

After you connect to your instance, you can verify that you see the NVMe volumes in Disk Manager. On the taskbar, open the context (right-click) menu for the Windows logo and choose **Disk Management**. On Windows Server 2008 R2, choose **Start**, **Administrative Tools**, **Computer Management**, **Disk Management**.

The AWS Windows AMIs provided by Amazon include the AWS NVMe driver. If you are not using the latest AWS Windows AMIs, you can install the current AWS NVMe driver (p. 446).

The data on NVMe instance storage is encrypted using an XTS-AES-256 block cipher implemented in a hardware module on the instance. The encryption keys are generated using the hardware module and are unique to each NVMe instance storage device. All encryption keys are destroyed when the instance is stopped or terminated and cannot be recovered. You cannot disable this encryption and you cannot provide your own encryption key.

**Instance Store Volume TRIM Support**

The following instances support SSD volumes with TRIM: C5d, F1, I2, I3, I3en, M5ad, M5d, p3dn.24xlarge, R3, R5ad, R5d, and z1d.

Instances running Windows Server 2012 R2 support TRIM as of AWS PV Driver version 7.3.0. Instances running earlier versions of Windows Server do not support TRIM.

Instance store volumes that support TRIM are fully trimmed before they are allocated to your instance. These volumes are not formatted with a file system when an instance launches, so you must format them before they can be mounted and used. For faster access to these volumes, you should skip the TRIM operation when you format them.

With instance store volumes that support TRIM, you can use the TRIM command to notify the SSD controller when you no longer need data that you’ve written. This provides the controller with more free space, which can reduce write amplification and increase performance. On Windows, use the `fsutil behavior set DisableDeleteNotify 1` command.

**File Storage**

Cloud file storage is a method for storing data in the cloud that provides servers and applications access to data through shared file systems. This compatibility makes cloud file storage ideal for workloads that rely on shared file systems and provides simple integration without code changes.

There are many file storage solutions that exist, ranging from a single node file server on a compute instance using block storage as the underpinnings with no scalability or few redundancies to protect the data, to a do-it-yourself clustered solution, to a fully-managed solution, such as Amazon Elastic File System (Amazon EFS) (p. 926) or Amazon FSx for Windows File Server (p. 926).
Amazon Elastic File System (Amazon EFS)

Amazon EFS provides scalable file storage for use with Amazon EC2. You can create an EFS file system and configure your instances to mount the file system. You can use an EFS file system as a common data source for workloads and applications running on multiple instances. For more information, see Amazon Elastic File System (Amazon EFS) in the Amazon EC2 User Guide for Linux Instances and the Amazon Elastic File System product page.

Important
Amazon EFS is not supported on Windows instances.

Amazon FSx for Windows File Server

Amazon FSx for Windows File Server provides fully managed Windows file servers, backed by a fully-native Windows file system with the features, performance, and compatibility to easily lift and shift enterprise applications to AWS.

Amazon FSx supports a broad set of enterprise Windows workloads with fully managed file storage built on Microsoft Windows Server. Amazon FSx has native support for Windows file system features and for the industry-standard Server Message Block (SMB) protocol to access file storage over a network. Amazon FSx is optimized for enterprise applications in the AWS Cloud, with native Windows compatibility, enterprise performance and features, and consistent sub-millisecond latencies.

With file storage on Amazon FSx, the code, applications, and tools that Windows developers and administrators use today can continue to work unchanged. The Windows applications and workloads that are ideal for Amazon FSx include business applications, home directories, web serving, content management, data analytics, software build setups, and media processing workloads.

As a fully managed service, Amazon FSx for Windows File Server eliminates the administrative overhead of setting up and provisioning file servers and storage volumes. Additionally, it keeps Windows software up to date, detects and addresses hardware failures, and performs backups. It also provides rich integration with other AWS services, including AWS Directory Service for Microsoft Active Directory, Amazon WorkSpaces, AWS Key Management Service, and AWS CloudTrail.

For more information, see the Amazon FSx for Windows File Server User Guide.

Amazon Simple Storage Service (Amazon S3)

Amazon S3 is a repository for Internet data. Amazon S3 provides access to reliable, fast, and inexpensive data storage infrastructure. It is designed to make web-scale computing easy by enabling you to store and retrieve any amount of data, at any time, from within Amazon EC2 or anywhere on the web. Amazon S3 stores data objects redundantly on multiple devices across multiple facilities and allows concurrent read or write access to these data objects by many separate clients or application threads. You can use the redundant data stored in Amazon S3 to recover quickly and reliably from instance or application failures.

Amazon EC2 uses Amazon S3 for storing Amazon Machine Images (AMIs). You use AMIs for launching EC2 instances. In case of instance failure, you can use the stored AMI to immediately launch another instance, thereby allowing for fast recovery and business continuity.

Amazon EC2 also uses Amazon S3 to store snapshots (backup copies) of the data volumes. You can use snapshots for recovering data quickly and reliably in case of application or system failures. You can also use snapshots as a baseline to create multiple new data volumes, expand the size of an existing data volume, or move data volumes across multiple Availability Zones, thereby making your data usage highly scalable. For more information about using data volumes and snapshots, see Amazon Elastic Block Store (p. 782).
Objects are the fundamental entities stored in Amazon S3. Every object stored in Amazon S3 is contained in a bucket. Buckets organize the Amazon S3 namespace at the highest level and identify the account responsible for that storage. Amazon S3 buckets are similar to Internet domain names. Objects stored in the buckets have a unique key value and are retrieved using a HTTP URL address. For example, if an object with a key value /photos/mygarden.jpg is stored in the myawsbucket bucket, then it is addressable using the URL http://myawsbucket.s3.amazonaws.com/photos/mygarden.jpg.

For more information about the features of Amazon S3, see the Amazon S3 product page.

Amazon S3 and Amazon EC2

Given the benefits of Amazon S3 for storage, you may decide to use this service to store files and data sets for use with EC2 instances. There are several ways to move data to and from Amazon S3 to your instances. In addition to the examples discussed below, there are a variety of tools that people have written that you can use to access your data in Amazon S3 from your computer or your instance. Some of the common ones are discussed in the AWS forums.

If you have permission, you can copy a file to or from Amazon S3 and your instance using one of the following methods.

AWS Tools for Windows PowerShell

Windows instances have the benefit of a graphical browser that you can use to access the Amazon S3 console directly; however, for scripting purposes, Windows users can also use the AWS Tools for Windows PowerShell to move objects to and from Amazon S3.

Use the following command to copy an Amazon S3 object to your Windows instance.

```
PS C:\> Copy-S3Object -BucketName my_bucket -Key path-to-file -LocalFile my_copied_file.ext
```

AWS Command Line Interface

The AWS Command Line Interface (AWS CLI) is a unified tool to manage your AWS services. The AWS CLI enables users to authenticate themselves and download restricted items from Amazon S3 and also to upload items. For more information, such as how to install and configure the tools, see the AWS Command Line Interface detail page.

The `aws s3 cp` command is similar to the Unix `cp` command. You can copy files from Amazon S3 to your instance, copy files from your instance to Amazon S3, and copy files from one Amazon S3 location to another.

Use the following command to copy an object from Amazon S3 to your instance.

```
aws s3 cp s3://my_bucket/my_folder/my_file.ext my_copied_file.ext
```

Use the following command to copy an object from your instance back into Amazon S3.

```
aws s3 cp my_copied_file.ext s3://my_bucket/my_folder/my_file.ext
```

The `aws s3 sync` command can synchronize an entire Amazon S3 bucket to a local directory location. This can be helpful for downloading a data set and keeping the local copy up-to-date with the remote set. If you have the proper permissions on the Amazon S3 bucket, you can push your local directory back up to the cloud when you are finished by reversing the source and destination locations in the command.

Use the following command to download an entire Amazon S3 bucket to a local directory on your instance.
AWS S3 API

If you are a developer, you can use an API to access data in Amazon S3. For more information, see the Amazon Simple Storage Service Developer Guide. You can use this API and its examples to help develop your application and integrate it with other APIs and SDKs, such as the boto Python interface.

Instance Volume Limits

The maximum number of volumes that your instance can have depends on the operating system and instance type. When considering how many volumes to add to your instance, you should consider whether you need increased I/O bandwidth or increased storage capacity.

Contents

- Linux-Specific Volume Limits (p. 928)
- Windows-Specific Volume Limits (p. 928)
- Instance Type Limits (p. 929)
- Bandwidth versus Capacity (p. 929)

Linux-Specific Volume Limits

Attaching more than 40 volumes can cause boot failures. Note that this number includes the root volume, plus any attached instance store volumes and EBS volumes. If you experience boot problems on an instance with a large number of volumes, stop the instance, detach any volumes that are not essential to the boot process, and then reattach the volumes after the instance is running.

Important
Attaching more than 40 volumes to a Linux instance is supported on a best effort basis only and is not guaranteed.

Windows-Specific Volume Limits

The following table shows the volume limits for Windows instances based on the driver used. Note that these numbers include the root volume, plus any attached instance store volumes and EBS volumes.

Important
Attaching more than the following volumes to a Windows instance is supported on a best effort basis only and is not guaranteed.

<table>
<thead>
<tr>
<th>Driver</th>
<th>Volume Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS PV</td>
<td>26</td>
</tr>
<tr>
<td>Citrix PV</td>
<td>26</td>
</tr>
<tr>
<td>Red Hat PV</td>
<td>17</td>
</tr>
</tbody>
</table>

We do not recommend that you give a Windows instance more than 26 volumes with AWS PV or Citrix PV drivers, as it is likely to cause performance issues.

To determine which PV drivers your instance is using, or to upgrade your Windows instance from Red Hat to Citrix PV drivers, see Upgrading PV Drivers on Your Windows Instances (p. 436).
Instance Type Limits

C5, C5d, C5n, G4, i3en, M5, M5a, M5ad, M5d, p3dn.24xlarge, R5, R5a, R5ad, R5d, T3, T3a, and z1d instances support a maximum of 28 attachments, including network interfaces, EBS volumes, and NVMe instance store volumes. Every instance has at least one network interface attachment. NVMe instance store volumes are automatically attached. For example, if you have no additional network interface attachments on an EBS-only instance, you can attach up to 27 EBS volumes to it. If you have one additional network interface on an instance with 2 NVMe instance store volumes, you can attach 24 EBS volumes to it. For more information, see Elastic Network Interfaces (p. 679) and Instance Store Volumes (p. 918).

i3.metal, m5.metal, m5d.metal, r5.metal, r5d.metal, and z1d.metal instances support a maximum of 31 EBS volumes.

u-6tb1.metal, u-9tb1.metal, and u-12tb1.metal instances support a maximum of 13 EBS volumes.

Bandwidth versus Capacity

For consistent and predictable bandwidth use cases, use EBS-optimized or 10 Gigabit network connectivity instances and General Purpose SSD or Provisioned IOPS SSD volumes. Follow the guidance in Amazon EC2 Instance Configuration (p. 888) to match the IOPS you have provisioned for your volumes to the bandwidth available from your instances for maximum performance. For RAID configurations, many administrators find that arrays larger than 8 volumes have diminished performance returns due to increased I/O overhead. Test your individual application performance and tune it as required.

Device Naming on Windows Instances

When you attach a volume to your instance, you include a device name for the volume. This device name is used by Amazon EC2. The block device driver for the instance assigns the actual volume name when mounting the volume, and the name assigned can be different from the name that Amazon EC2 uses.

The number of volumes that your instance can support is determined by the operating system. For more information, see Instance Volume Limits (p. 928).

Contents
- Available Device Names (p. 929)
- Device Name Considerations (p. 930)

For information about device names on Linux instances, see Device Naming on Linux Instances in the Amazon EC2 User Guide for Linux Instances.

Available Device Names

Windows AMIs use one of the following sets of drivers to permit access to virtualized hardware: AWS PV, Citrix PV, and RedHat PV. For more information, see Paravirtual Drivers for Windows Instances (p. 431).

The following table lists the available device names that you can specify in a block device mapping or when attaching an EBS volume.
Device Name Considerations

Keep the following in mind when selecting a device name:

- Although you can attach your EBS volumes using the device names used to attach instance store volumes, we strongly recommend that you don’t because the behavior can be unpredictable.
- The number of NVMe instance store volumes for an instance depends on the size of the instance. NVMe instance store volumes are automatically enumerated and assigned a Windows drive letter.
- AWS Windows AMIs come with additional software that prepares an instance when it first boots up. This is either the EC2Config service (Windows AMIs prior to Windows Server 2016) or EC2Launch (Windows Server 2016 and later). After the devices have been mapped to drives, they are initialized and mounted. The root drive is initialized and mounted as C:\. The instance store volumes attached to the instance are initialized and mounted as Z:\, Y:\, and so on. By default, when an EBS volume is attached to a Windows instance, it can show up as any drive letter on the instance. You can change the settings to set the drive letters of the volumes per your specifications. For more information, see Configuring a Windows Instance Using the EC2Config Service (p. 405), Configuring a Windows Instance Using EC2Launch (p. 395), and Mapping Disks to Volumes on Your Windows Instance (p. 939).

Block Device Mapping

Each instance that you launch has an associated root device volume, either an Amazon EBS volume or an instance store volume. You can use block device mapping to specify additional EBS volumes or instance store volumes to attach to an instance when it’s launched. You can also attach additional EBS volumes to a running instance; see Attaching an Amazon EBS Volume to an Instance (p. 803). However, the only way to attach instance store volumes to an instance is to use block device mapping to attach them as the instance is launched.

For more information about root device volumes, see Root Device Volume (p. 7).
Block Device Mapping Concepts

A block device is a storage device that moves data in sequences of bytes or bits (blocks). These devices support random access and generally use buffered I/O. Examples include hard disks, CD-ROM drives, and flash drives. A block device can be physically attached to a computer or accessed remotely as if it were physically attached to the computer. Amazon EC2 supports two types of block devices:

- Instance store volumes (virtual devices whose underlying hardware is physically attached to the host computer for the instance)
- EBS volumes (remote storage devices)

A block device mapping defines the block devices (instance store volumes and EBS volumes) to attach to an instance. You can specify a block device mapping as part of creating an AMI so that the mapping is used by all instances launched from the AMI. Alternatively, you can specify a block device mapping when you launch an instance, so this mapping overrides the one specified in the AMI from which you launched the instance. Note that all NVMe instance store volumes supported by an instance type are automatically enumerated and assigned a device name on instance launch; including them in your block device mapping has no effect.

Block Device Mapping Entries

When you create a block device mapping, you specify the following information for each block device that you need to attach to the instance:

- The device name used within Amazon EC2. The block device driver for the instance assigns the actual volume name when mounting the volume. The name assigned can be different from the name that Amazon EC2 recommends. For more information, see Device Naming on Windows Instances (p. 929).
- [Instance store volumes] The virtual device: ephemeral[0–23]. Note that the number and size of available instance store volumes for your instance varies by instance type.
- [NVMe instance store volumes] These volumes are automatically enumerated and assigned a device name; including them in your block device mapping has no effect.
- [EBS volumes] The ID of the snapshot to use to create the block device (snap-xxxxxxxx). This value is optional as long as you specify a volume size.
- [EBS volumes] The size of the volume, in GiB. The specified size must be greater than or equal to the size of the specified snapshot.
- [EBS volumes] Whether to delete the volume on instance termination (true or false). The default value is true for the root device volume and false for attached volumes. When you create an AMI, its block device mapping inherits this setting from the instance. When you launch an instance, it inherits this setting from the AMI.
• [EBS volumes] The volume type, which can be gp2 for General Purpose SSD, io1 for Provisioned IOPS SSD, st1 for Throughput Optimized HDD, sc1 for Cold SSD, or standard for Magnetic. The default value is gp2.

• [EBS volumes] The number of input/output operations per second (IOPS) that the volume supports. (Not used with gp2, st1, sc1, or standard volumes.)

Block Device Mapping Instance Store Caveats

There are several caveats to consider when launching instances with AMIs that have instance store volumes in their block device mappings.

• Some instance types include more instance store volumes than others, and some instance types contain no instance store volumes at all. If your instance type supports one instance store volume, and your AMI has mappings for two instance store volumes, then the instance launches with one instance store volume.

• Instance store volumes can only be mapped at launch time. You cannot stop an instance without instance store volumes (such as the t2.micro), change the instance to a type that supports instance store volumes, and then restart the instance with instance store volumes. However, you can create an AMI from the instance and launch it on an instance type that supports instance store volumes, and map those instance store volumes to the instance.

• If you launch an instance with instance store volumes mapped, and then stop the instance and change it to an instance type with fewer instance store volumes and restart it, the instance store volume mappings from the initial launch still show up in the instance metadata. However, only the maximum number of supported instance store volumes for that instance type are available to the instance.

  Note
  When an instance is stopped, all data on the instance store volumes is lost.

• Depending on instance store capacity at launch time, M3 instances may ignore AMI instance store block device mappings at launch unless they are specified at launch. You should specify instance store block device mappings at launch time, even if the AMI you are launching has the instance store volumes mapped in the AMI, to ensure that the instance store volumes are available when the instance launches.

Example Block Device Mapping

This figure shows an example block device mapping for an EBS-backed instance. It maps /dev/sdb to ephemeral0 and maps two EBS volumes, one to /dev/sdh and the other to /dev/sdj. It also shows the EBS volume that is the root device volume, /dev/sda1.
How Devices Are Made Available in the Operating System

Device names like `/dev/sdh` and `xvdh` are used by Amazon EC2 to describe block devices. The block device mapping is used by Amazon EC2 to specify the block devices to attach to an EC2 instance. After a block device is attached to an instance, it must be mounted by the operating system before you can access the storage device. When a block device is detached from an instance, it is unmounted by the operating system and you can no longer access the storage device.

With a Windows instance, the device names specified in the block device mapping are mapped to their corresponding block devices when the instance first boots, and then the Ec2Config service initializes and mounts the drives. The root device volume is mounted as `C:\`. The instance store volumes are mounted as `Z:\`, `Y:\`, and so on. When an EBS volume is mounted, it can be mounted using any available drive letter. However, you can configure how the Ec2Config Service assigns drive letters to EBS volumes; for more information, see Configuring a Windows Instance Using the EC2Config Service (p. 405).

AMI Block Device Mapping

Each AMI has a block device mapping that specifies the block devices to attach to an instance when it is launched from the AMI. An AMI that Amazon provides includes a root device only. To add more block devices to an AMI, you must create your own AMI.
Specifying a Block Device Mapping for an AMI

There are two ways to specify volumes in addition to the root volume when you create an AMI. If you've already attached volumes to a running instance before you create an AMI from the instance, the block device mapping for the AMI includes those same volumes. For EBS volumes, the existing data is saved to a new snapshot, and it's this new snapshot that's specified in the block device mapping. For instance store volumes, the data is not preserved.

For an EBS-backed AMI, you can add EBS volumes and instance store volumes using a block device mapping. For an instance store-backed AMI, you can add instance store volumes only by modifying the block device mapping entries in the image manifest file when registering the image.

Note
For M3 instances, you must specify instance store volumes in the block device mapping for the instance when you launch it. When you launch an M3 instance, instance store volumes specified in the block device mapping for the AMI may be ignored if they are not specified as part of the instance block device mapping.

To add volumes to an AMI using the console

1. Open the Amazon EC2 console.
2. In the navigation pane, choose Instances.
4. In the Create Image dialog box, choose Add New Volume.
5. Select a volume type from the Type list and a device name from the Device list. For an EBS volume, you can optionally specify a snapshot, volume size, and volume type.
6. Choose Create Image.

To add volumes to an AMI using the command line

Use the create-image AWS CLI command to specify a block device mapping for an EBS-backed AMI. Use the register-image AWS CLI command to specify a block device mapping for an instance store-backed AMI.

Specify the block device mapping using the --block-device-mappings parameter. Arguments encoded in JSON can be supplied either directly on the command line or by reference to a file:

```
--block-device-mappings [mapping, ...]
--block-device-mappings [file://mapping.json]
```

To add an instance store volume, use the following mapping:

```
{
    "DeviceName": "xvdb",
    "VirtualName": "ephemeral0"
}
```

To add an empty 100 GiB gp2 volume, use the following mapping:

```
{
    "DeviceName": "xvdg",
    "Ebs": {
        "VolumeSize": 100
    }
}
```
To add an EBS volume based on a snapshot, use the following mapping:

```json
{  
    "DeviceName": "xvdh",
    "Ebs": {  
        "SnapshotId": "snap-xxxxxxxx"
    }
}
```

To omit a mapping for a device, use the following mapping:

```json
{  
    "DeviceName": "xvdj",
    "NoDevice": ""
}
```

Alternatively, you can use the `-BlockDeviceMapping` parameter with the following commands (AWS Tools for Windows PowerShell):

- New-EC2Image
- Register-EC2Image

### Viewing the EBS Volumes in an AMI Block Device Mapping

You can easily enumerate the EBS volumes in the block device mapping for an AMI.

#### To view the EBS volumes for an AMI using the console

1. Open the Amazon EC2 console.
2. In the navigation pane, choose AMIs.
3. Choose EBS images from the Filter list to get a list of EBS-backed AMIs.
4. Select the desired AMI, and look at the Details tab. At a minimum, the following information is available for the root device:

   - **Root Device Type** (ebs)
   - **Root Device Name** (for example, /dev/sda1)
   - **Block Devices** (for example, /dev/sda1=snap-1234567890abcdef0:8:true)

   If the AMI was created with additional EBS volumes using a block device mapping, the Block Devices field displays the mapping for those additional volumes as well. (Recall that this screen doesn’t display instance store volumes.)

#### To view the EBS volumes for an AMI using the command line

Use the `describe-images` (AWS CLI) command or `Get-EC2Image` (AWS Tools for Windows PowerShell) command to enumerate the EBS volumes in the block device mapping for an AMI.

### Instance Block Device Mapping

By default, an instance that you launch includes any storage devices specified in the block device mapping of the AMI from which you launched the instance. You can specify changes to the block device mapping for an instance when you launch it, and these updates overwrite or merge with the block device mapping of the AMI.
Limits

- For the root volume, you can only modify the following: volume size, volume type, and the **Delete on Termination** flag.
- When you modify an EBS volume, you can't decrease its size. Therefore, you must specify a snapshot whose size is equal to or greater than the size of the snapshot specified in the block device mapping of the AMI.

Contents

- Updating the Block Device Mapping when Launching an Instance (p. 936)
- Updating the Block Device Mapping of a Running Instance (p. 937)
- Viewing the EBS Volumes in an Instance Block Device Mapping (p. 938)
- Viewing the Instance Block Device Mapping for Instance Store Volumes (p. 938)

Updating the Block Device Mapping when Launching an Instance

You can add EBS volumes and instance store volumes to an instance when you launch it. Note that updating the block device mapping for an instance doesn't make a permanent change to the block device mapping of the AMI from which it was launched.

To add volumes to an instance using the console

1. Open the Amazon EC2 console.
2. From the dashboard, choose **Launch Instance**.
3. On the **Choose an Amazon Machine Image (AMI)** page, select the AMI to use and choose **Select**.
4. Follow the wizard to complete the **Choose an Instance Type** and **Configure Instance Details** pages.
5. On the **Add Storage** page, you can modify the root volume, EBS volumes, and instance store volumes as follows:
   - To change the size of the root volume, locate the **Root** volume under the **Type** column, and change its **Size** field.
   - To suppress an EBS volume specified by the block device mapping of the AMI used to launch the instance, locate the volume and click its **Delete** icon.
   - To add an EBS volume, choose **Add New Volume**, choose **EBS** from the **Type** list, and fill in the fields (**Device**, **Snapshot**, and so on).
   - To suppress an instance store volume specified by the block device mapping of the AMI used to launch the instance, locate the volume, and choose its **Delete** icon.
   - To add an instance store volume, choose **Add New Volume**, select **Instance Store** from the **Type** list, and select a device name from **Device**.
6. Complete the remaining wizard pages, and choose **Launch**.

To add volumes to an instance using the command line

Use the **run-instances** AWS CLI command to specify a block device mapping for an instance.

Specify the block device mapping using the following parameter:

```
--block-device-mappings [mapping, ...]
```

For example, suppose that an EBS-backed AMI specifies the following block device mapping:
• xvdb=ephemeral0
• xvdh=snap-1234567890abcdef0
• xvdj=:100

To prevent xvdj from attaching to an instance launched from this AMI, use the following mapping:

```json
{
    "DeviceName": "xvdj",
    "NoDevice": ""
}
```

To increase the size of xvdh to 300 GiB, specify the following mapping. Notice that you don’t need to specify the snapshot ID for xvdh, because specifying the device name is enough to identify the volume.

```json
{
    "DeviceName": "xvdh",
    "Ebs": {
        "VolumeSize": 300
    }
}
```

To attach an additional instance store volume, xvdc, specify the following mapping. If the instance type doesn’t support multiple instance store volumes, this mapping has no effect.

```json
{
    "DeviceName": "xvdc",
    "VirtualName": "ephemeral1"
}
```

Alternatively, you can use the -BlockDeviceMapping parameter with the New-EC2Instance command (AWS Tools for Windows PowerShell).

### Updating the Block Device Mapping of a Running Instance

You can use the following modify-instance-attribute AWS CLI command to update the block device mapping of a running instance. Note that you do not need to stop the instance before changing this attribute.

```
aws ec2 modify-instance-attribute --instance-id i-1a2b3c4d --block-device-mappings file://mapping.json
```

For example, to preserve the root volume at instance termination, specify the following in mapping.json:

```json
[
    {
        "DeviceName": "/dev/sda1",
        "Ebs": {
            "DeleteOnTermination": false
        }
    }
]
```

Alternatively, you can use the -BlockDeviceMapping parameter with the Edit-EC2InstanceAttribute command (AWS Tools for Windows PowerShell).
Viewing the EBS Volumes in an Instance Block Device Mapping

You can easily enumerate the EBS volumes mapped to an instance.

**Note**
For instances launched before the release of the 2009-10-31 API, AWS can't display the block device mapping. You must detach and reattach the volumes so that AWS can display the block device mapping.

**To view the EBS volumes for an instance using the console**

1. Open the Amazon EC2 console.
2. In the navigation pane, choose **Instances**.
3. In the search bar, type **Root Device Type**, and then choose **EBS**. This displays a list of EBS-backed instances.
4. Select the desired instance and look at the details displayed in the **Description** tab. At a minimum, the following information is available for the root device:

   - **Root device type** (ebs)
   - **Root device** (for example, `/dev/sda1`)
   - **Block devices** (for example, `/dev/sda1`, `xvdh`, and `xvdj`)

   If the instance was launched with additional EBS volumes using a block device mapping, the **Block devices** field displays those additional volumes as well as the root device. (Recall that this dialog box doesn't display instance store volumes.)

5. To display additional information about a block device, select its entry next to **Block devices**. This displays the following information for the block device:

   - **EBS ID** (vol-xxxxxxxx)
   - **Root device type** (ebs)
   - **Attachment time** (yyyy-mmThh:mm:ssZ)
   - **Block device status** (attaching, attached, detaching, detached)
   - **Delete on termination** (Yes, No)

**To view the EBS volumes for an instance using the command line**

Use the `describe-instances` (AWS CLI) command or `Get-EC2Instance` (AWS Tools for Windows PowerShell) command to enumerate the EBS volumes in the block device mapping for an instance.

Viewing the Instance Block Device Mapping for Instance Store Volumes

When you view the block device mapping for your instance, you can see only the EBS volumes, not the instance store volumes. You can use instance metadata to query the complete block device mapping. The base URI for all requests for instance metadata is `http://169.254.169.254/latest/`.

**Important**
NVMe instance store volumes are not included in the block device mapping.
First, connect to your running instance. From the instance, use this query to get its block device mapping.

```powershell
```

The response includes the names of the block devices for the instance. For example, the output for an instance store–backed `m1.small` instance looks like this:

```
ami
ephemeral0
root
swap
```

The `ami` device is the root device as seen by the instance. The instance store volumes are named `ephemeral[0-23]`. The `swap` device is for the page file. If you’ve also mapped EBS volumes, they appear as `ebs1`, `ebs2`, and so on.

To get details about an individual block device in the block device mapping, append its name to the previous query, as shown here.

```powershell
```

For more information, see Instance Metadata and User Data (p. 478).

### Mapping Disks to Volumes on Your Windows Instance

Your Windows instance comes with an EBS volume that serves as the root volume. If your Windows instance uses AWS PV or Citrix PV drivers, you can optionally add up to 25 volumes, making a total of 26 volumes. For more information, see Instance Volume Limits (p. 928)

Depending on the instance type of your instance, you’ll have from 0 to 24 possible instance store volumes available to the instance. To use any of the instance store volumes that are available to your instance, you must specify them when you create your AMI or launch your instance. You can also add EBS volumes when you create your AMI or launch your instance, or attach them while your instance is running. For more information, see Making an Amazon EBS Volume Available for Use on Windows (p. 804).

When you add a volume to your instance, you specify the device name that Amazon EC2 uses. For more information, see Device Naming on Windows Instances (p. 929). AWS Windows Amazon Machine Images (AMIs) contain a set of drivers that are used by Amazon EC2 to map instance store and EBS volumes to Windows disks and drive letters. If you launch an instance from a Windows AMI that uses AWS PV or Citrix PV drivers, you can use the relationships described on this page to map your Windows disks to your instance store and EBS volumes. If your Windows AMI uses Red Hat PV drivers, you can update your instance to use the Citrix drivers. For more information, see Upgrading PV Drivers on Your Windows Instances (p. 436).

**Contents**

- Listing the Disks Using Windows Disk Management (p. 940)
- Listing the Disks Using Windows PowerShell (Windows Server 2012 and later) (p. 941)
- Disk Device to Device Name Mapping (p. 943)
Listing the Disks Using Windows Disk Management

You can find the disks on your Windows instance using Windows Disk Management.

To find the disks on your Windows instance

1. Log in to your Windows instance using Remote Desktop. For more information, see, Connecting to Your Windows Instance (p. 379).
2. Start the Disk Management utility.
   
3. Review the disks. The root volume is an EBS volume mounted as C:\. If there are no other disks shown, then you didn’t specify additional volumes when you created the AMI or launched the instance.

The following is an example that shows the disks that are available if you launch an m3.medium instance with an instance store volume (Disk 2) and an additional EBS volume (Disk 1).

4. Right-click the gray pane labeled Disk 1, and then select Properties. Note the value of Location and look it up in the tables in Disk Device to Device Name Mapping (p. 943). For example, the following disk has the location Bus Number 0, Target Id 9, LUN 0. According to the table for EBS volumes, the device name for this location is xvdj.
5. To map the device name of an EBS volume to its volume ID, open the Amazon EC2 console on your computer. In the navigation pane, select **Instances**, and then select your instance. Under **Block devices** on the **Description** tab, click the device name, and locate **EBS ID**. For this example, the volume ID is `vol-0a07f3e37b14708b9`.

Note that the Amazon EC2 console shows only the EBS volumes.

Another way to access volume IDs from the console is to select **Volumes** under **Elastic Block Store** in the navigation pane. Volumes will be listed by instance name and volume size. Review or select the **Attachment Information** to verify the instance to which the volume is attached.

### Listing the Disks Using Windows PowerShell
(Windows Server 2012 and later)

The following PowerShell script lists each disk and its corresponding device name and volume.

```powershell
# List the Windows disks
function Get-EC2InstanceMetadata {
    param([string]$Path)
    (Invoke-WebRequest -Uri "http://169.254.169.254/latest/$Path").Content
}
```
function Convert-SCSITargetIdToDeviceName {
    param([int]$SCSITargetId)
    If ($SCSITargetId -eq 0) {
        return "sda1"
    }
    $deviceName = "xvd"
    If ($SCSITargetId -gt 25) {
        $deviceName += [char](0x60 + [int]($SCSITargetId / 26))
    }
    $deviceName += [char](0x61 + $SCSITargetId % 26)
    return $deviceName
}

Try {
    $InstanceId = Get-EC2InstanceMetadata "meta-data/instance-id"
    $AZ = Get-EC2InstanceMetadata "meta-data/placement/availability-zone"
    $Region = $AZ.Remove($AZ.Length - 1)
    $BlockDeviceMappings = (Get-EC2Instance -Region $Region -Instance $InstanceId).Instances.BlockDeviceMappings
    $VirtualDeviceMap = @{
        (Get-EC2InstanceMetadata "meta-data/block-device-mapping").Split("\n") | ForEach-Object {
            $VirtualDevice = $_
            $BlockDeviceName = Get-EC2InstanceMetadata "meta-data/block-device-mapping/
            $VirtualDevice"
            $VirtualDeviceMap[$BlockDeviceName] = $VirtualDevice
        }
    }
}
Catch {
    Write-Host "Could not access the AWS API, therefore, VolumeId is not available. Verify that you provided your access keys."
}

Get-disk | ForEach-Object {
    $DriveLetter = $null
    $VolumeName = $null
    $DiskDrive = $_
    $Disk = $_.Number
    $Partitions = $_.NumberOfPartitions
    $EbsVolumeID = $_.SerialNumber -replace "_\[^\s\]*$" -replace "vol", "vol-"
    Get-Partition -DiskId $_.Path | ForEach-Object {
        if ($_.DriveLetter -ne """) {
            $DriveLetter = $_.DriveLetter
            $VolumeName = (Get-PSDrive | Where-Object {$_.Name -eq $DriveLetter}).Description
        }
    }

    If (($DiskDrive.path -like "*PROD_PVDISK*")) {
        $BlockDeviceName = Convert-SCSITargetIdToDeviceName($Get-WmiObject -Class Win32_Diskdrive | Where-Object {$_.DeviceID -eq "$\\.\PHYSICALDRIVE\$DiskDrive.Number"}).SCSITargetId
        $BlockDeviceName = "/dev/" + $BlockDeviceName
        $BlockDevice = $BlockDeviceMappings | Where-Object { $BlockDeviceName -like "*" + $_.DeviceName "*" }
        $EbsVolumeID = $BlockDevice.Ebs.VolumeId
        $VirtualDevice = If ($VirtualDeviceMap.ContainsKey($BlockDeviceName)) {
            $VirtualDeviceMap[$BlockDeviceName]
        } Else { $null }
    }ElseIf (($DiskDrive.path -like "*PROD_AMAZON_EC2_NVME*")) {
        $BlockDeviceName = Get-EC2InstanceMetadata "meta-data/block-device-mapping/ephemeral"($Get-WmiObject -Class Win32_Diskdrive | Where-Object {$_.DeviceID -eq "$\\.\PHYSICALDRIVE\$DiskDrive.Number"}).SCSIPort - 2"
        $BlockDevice = $null
    Else {
        $null
    }
}
# Disk Device to Device Name Mapping

The block device driver for the instance assigns the actual volume names when mounting volumes.

## Mappings

- **Instance Store Volumes (p. 944)**
- **EBS Volumes (p. 944)**
- **NVMe EBS Volumes (p. 945)**
## Instance Store Volumes

The following table describes how the Citrix PV and AWS PV drivers map non-NVMe instance store volumes to Windows volumes. The number of available instance store volumes is determined by the instance type. For more information, see Instance Store Volumes (p. 918).

<table>
<thead>
<tr>
<th>Location</th>
<th>Device Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus Number 0, Target ID 78, LUN 0</td>
<td>xvdca</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 79, LUN 0</td>
<td>xvdcb</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 80, LUN 0</td>
<td>xvdcc</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 81, LUN 0</td>
<td>xvdcd</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 82, LUN 0</td>
<td>xvdce</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 83, LUN 0</td>
<td>xvdcf</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 84, LUN 0</td>
<td>xvdcg</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 85, LUN 0</td>
<td>xvdch</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 86, LUN 0</td>
<td>xvdci</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 87, LUN 0</td>
<td>xvdcj</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 88, LUN 0</td>
<td>xvdck</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 89, LUN 0</td>
<td>xvdcl</td>
</tr>
</tbody>
</table>

## EBS Volumes

The following table describes how the Citrix PV and AWS PV drivers map non-NVME EBS volumes to Windows volumes.

<table>
<thead>
<tr>
<th>Location</th>
<th>Device Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus Number 0, Target ID 0, LUN 0</td>
<td>/dev/sda1</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 1, LUN 0</td>
<td>xvdb</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 2, LUN 0</td>
<td>xvdc</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 3, LUN 0</td>
<td>xvd3</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 4, LUN 0</td>
<td>xvde</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 5, LUN 0</td>
<td>xvdf</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 6, LUN 0</td>
<td>xvdg</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 7, LUN 0</td>
<td>xvdh</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 8, LUN 0</td>
<td>xvd8</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 9, LUN 0</td>
<td>xvdj</td>
</tr>
<tr>
<td>Location</td>
<td>Device Name</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 10, LUN 0</td>
<td>xvdk</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 11, LUN 0</td>
<td>xvdl</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 12, LUN 0</td>
<td>xvdm</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 13, LUN 0</td>
<td>xvdn</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 14, LUN 0</td>
<td>xvdo</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 15, LUN 0</td>
<td>xvdp</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 16, LUN 0</td>
<td>xvdq</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 17, LUN 0</td>
<td>xvdr</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 18, LUN 0</td>
<td>xvds</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 19, LUN 0</td>
<td>xvdt</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 20, LUN 0</td>
<td>xvdu</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 21, LUN 0</td>
<td>xvdv</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 22, LUN 0</td>
<td>xvdw</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 23, LUN 0</td>
<td>xvdx</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 24, LUN 0</td>
<td>xvdy</td>
</tr>
<tr>
<td>Bus Number 0, Target ID 25, LUN 0</td>
<td>xvdz</td>
</tr>
</tbody>
</table>

**NVMe EBS Volumes**

With C5 and M5 instances, EBS volumes are exposed as NVMe devices. You can use the `Get-Disk` command to map Windows disk numbers to EBS volume IDs. For more information, see Identifying the EBS Device (p. 875).

**Tutorial: Deploy Storage Spaces Direct (S2D) on Amazon EC2**

Storage Spaces Direct (S2D) is a highly-scalable, software-defined storage architecture that enables users to cluster local storage with features in Windows Server 2016. S2D is an alternative to traditional SAN or NAS arrays. It uses built-in Windows features and tools to configure highly-available storage that crosses multiple nodes in a cluster. For more information, see Storage Spaces Direct in the Microsoft documentation.

The following diagram shows the architecture of S2D on Amazon EC2 Windows.
Skill Level

A basic understanding of Windows Server computing as well as how to create and manage domain-joined Amazon EC2 Windows instances in a VPC is required. Knowledge of the AWS Tools for Windows PowerShell and Windows Failover Clustering is helpful, but not required.

What you will accomplish in this tutorial

- Provision a highly-available storage cluster using Storage Spaces Direct (S2D).
- Provision a fault-tolerant, cluster-shared volume (CSV) on your cluster.

Before you begin

- If you haven't done so already, open https://aws.amazon.com/ and create an AWS account.
- Create a virtual private cloud (VPC) with a public subnet and two private subnets for your instances. A third, private, subnet should be configured for AWS Directory Service.
- Select one of the latest Amazon Machine Images (AMIs) for Windows Server 2016. You can use this AMI as is, or use it as the basis for your own custom AMI. AWS recommends using the latest public EC2 Windows Server 2016 AMI.
- Create an AWS Directory Service directory. This is no longer a requirement for enabling the Failover Clustering feature in Windows Server 2016. However, this tutorial assumes that your instances will be joined to an Active Directory domain, either on EC2 or AWS Managed Active Directory. For more information, see Getting Started with AWS Directory Service in the AWS Directory Service Administration Guide.
- Install and configure the AWS Tools for Windows PowerShell on your computer. For more information, see the AWS Tools for Windows PowerShell User Guide.

Important considerations

- Stopping instances with Instance Store Volumes (p. 917) can cause data loss if the data is not backed up or replicated. The data in an instance store persists only during the lifetime of its associated instance. If an instance reboots (intentionally or unintentionally), data in the instance store persists. However, data in the instance store is lost under the following circumstances:
  - The underlying disk drive fails.
  - The instance stops.
  - The instance terminates.

- Stopping too many instances in a cluster can cause data loss if the data is not backed up or replicated. When you use S2D on AWS, as with any cluster, losing more nodes than your fault tolerance allows will result in loss of data. One of the biggest risks to any cluster is losing all nodes. Cluster redundancy protects against failures on a single instance (or more, if your fault tolerance supports it). However,
Step 1: Launch and Domain Join Instances

All Nitro instances support Storage Spaces Direct using EBS and/or NVMe. All current generation Xen-based instances support Storage Spaces Direct with installation of AWS PV driver 8.2.3 and later. The best performance for storage can be achieved using i3 instances because they provide local instance store with NVMe and high network performance. Configuring S2D on Amazon EC2 requires a cluster of
at least two, but no more than 16 instances. These instances must each have at least two NVMe devices with high performance network connections between nodes, and run Windows Server 2016. For more information, see Storage Spaces Direct hardware requirements in the Microsoft documentation.

We recommend the i3 instance size because it satisfies the S2D hardware requirements and includes the largest and fastest instance store devices available. It also includes enhanced networking, which maximizes the available resources for S2D per instance. You can use M5D and R5D instance types, which have at least 2 NVMe disks, but local instance store disks will be used as cache disks for the storage spaces direct cluster and at least 2 EBS volumes will have to be added to each instance to provide capacity storage.

We recommend that you launch three instances to take advantage of three-way mirroring S2D fault tolerance, which enables you to conduct maintenance on a single node while maintaining fault tolerance in your cluster if a witness such as a file share witness is configured. You can also use two-way mirroring with two instances as a less expensive solution, but a witness will be necessary and high availability will not be maintained during maintenance on a cluster node.

We will deploy a two node cluster architecture using a file share witness hosted on an existing bastion machine that acts as our administration workstation. Each cluster node must be deployed in a different subnet. This architecture will be deployed into a single availability zone because Microsoft does not currently support stretch cluster with Storage Spaces Direct. However, the performance of a single availability zone and multi-availability zones are exactly the same as a result of our very low-latency and high-bandwidth design for availability zones.

To launch instances for your cluster

1. Using the Amazon EC2 console or the New-EC2Instance cmdlet, launch two i3.8xlarge instances to create the cluster and a t2.medium instance as an administration workstation and to host the file share witness. Use a different subnet for each instance. If you wish to follow a logic for IP assignment, then define the primary private IP address at creation time. In this case, you will need to define a secondary private IP address for each cluster node because the secondary IP will be assigned to the cluster VIP later.

To create each instance with PowerShell, use the New-EC2Instance command.

```
New-EC2Instance -ImageId ami-c49c0dac -MinCount 1 -MaxCount 1 -KeyName myPSKeyPair -SecurityGroupId mySGID -InstanceType i3.8xlarge -SubnetId mysubnetID
```

To create an AWS AD directory with PowerShell, use the New-DSMicrosoftAD command (or, refer to Create Your AWS Managed Microsoft AD Directory in AWS).

```
New-DSMicrosoftAD -Name corp.example.com –ShortName corp –Password P@ssw0rd –Description “AWS DS Managed” –VpcSettings_VpcId vpc-xxxxxxxx –VpcSettings_SubnetId subnet-xxxxxxxx, subnet-xxxxxxx
```

We use the following S2D-node1 network interface configuration:

```
Network interfaces

<table>
<thead>
<tr>
<th>Device</th>
<th>Network Interface</th>
<th>Subnet</th>
<th>Primary IP</th>
<th>Secondary IP addresses</th>
<th>IPv4 IPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>eth0</td>
<td>New network interface</td>
<td>subnet-9850a0fe</td>
<td>172.16.1.109</td>
<td>172.16.1.200</td>
<td>Add IP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>172.16.1.201</td>
<td></td>
<td>Add IP</td>
</tr>
</tbody>
</table>
```

Note
Each role deployed on this cluster, such as a SQL Failover Cluster instance or file server, will require additional secondary IP addresses on each node. The exception is the Scale-Out File Server role, which does not require an access point.
We use the following configuration:

<table>
<thead>
<tr>
<th>Server NetBIOS Name</th>
<th>IP Address</th>
<th>Subnets</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2D-Node1</td>
<td>172.16.1.199 (Primary)</td>
<td>AZ1 (e.g., eu-west-1a)</td>
</tr>
<tr>
<td></td>
<td>172.16.1.200 (secondary which will used for the cluster VIP)</td>
<td>– private subnet 1</td>
</tr>
<tr>
<td></td>
<td>172.16.1.201 (secondary which will be used later for a role such as SQL FCI)</td>
<td></td>
</tr>
<tr>
<td>S2D-Node2</td>
<td>172.16.3.199 (Primary)</td>
<td>AZ1 (e.g., eu-west-1a)</td>
</tr>
<tr>
<td></td>
<td>172.16.3.200 (secondary which will be used for the cluster VIP)</td>
<td>– private subnet 2</td>
</tr>
<tr>
<td></td>
<td>172.16.3.201 (secondary which will be used later for a role such as SQL FCI)</td>
<td></td>
</tr>
<tr>
<td>ADM01</td>
<td>Not specified</td>
<td>AZ1 (e.g., eu-west-1a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– public subnet</td>
</tr>
</tbody>
</table>

2. You can use seamless domain join at creation time to join instances to the domain. If you want to join them to the domain after they are launched, use the Add-Computer command. We recommend using AWS Systems Manager and AWS Directory Service to seamlessly join EC2 instances to a domain.

The steps in the remainder of this tutorial require execution with a domain account with local administrative privileges on each instance. Rename the instances as you want them before moving to the configuration. Ensure that your security groups and Windows firewalls are properly configured to allow remote PowerShell connection and cluster communications on these nodes.

**Step 2: Install and Configure Instance Prerequisites**

S2D requires File Services and Failover Clustering Window features, and at least one ten Gbps network interface. We recommend that you configure SMB to use SMB Multichannel, with RSS client connection counts that match the RSS queue count of the enhanced network adapter.

The following steps will be accomplished from the bastion instance ADM01.

**To install required Windows features**

- Install the File Services and Failover-Clustering Windows features with the management tools on cluster nodes. Install only failover management tools on ADM01.

  **Note**
  Change "S2D-Node1" and "S2D-Node2" to reflect the computer names that you set for the two instances; otherwise, the values will not change.
To configure networking

1. Enable multichannel and set the RSS Connection Count.

```powershell
foreach ($node in $nodes) {
    Invoke-Command -ComputerName $node -ScriptBlock {
        [int]$RssQCount = (Get-NetAdapterAdvancedProperty | Where DisplayName -like "Maximum Number of RSS Queues").RegistryValue | Select -First 1
        $Params = @{
            EnableMultiChannel                    = $true;
            ConnectionCountPerRssNetworkInterface = $RssQCount;
            Confirm                               = $false;
        }
        Set-SmbClientConfiguration @Params
    }
}
```

2. Configure RSS.

```powershell
foreach ($node in $nodes) {
    Invoke-Command -ComputerName $node -ScriptBlock {
        Get-WmiObject -class Win32_processor | ft systemname, Name, DeviceID,
        NumberOfCores, NumberOfLogicalProcessors
        $maxvcpu = (Get-WmiObject -class Win32_processor).NumberOfLogicalProcessors
        Get-NetAdapter | Set-NetAdapterRss -BaseProcessorNumber 2 -MaxProcessors $maxvcpu
    }
}
```

**Note**
You will see a disconnection message when executing this command because the network adapter restarts after setting the RSS configuration.

Receive Side Scaling (RSS) is a very important technology in networking on Windows. RSS ensures that incoming network traffic is spread among the available processors in the server for processing. If RSS is not used, network processing is bound to one processor, which is limited to approximately 4GBps. Currently, every NIC, by default, enables RSS, but the configuration is not optimized. Every NIC is configured, by default, with "Base Processor" 0, which means it will start processing on processor 0 together with the others NICs. To optimally configure RSS, start at processor 1 so we don't interfere with processes landing default on processor 0.

3. Increase storage space I/O timeout value to 30 seconds (recommended when configured into a guest cluster).

```powershell
foreach ($node in $nodes) {
    Invoke-Command -ComputerName $node -ScriptBlock {
        Set-ItemProperty -Path HKLM:\SYSTEM\CurrentControlSet\Services\spaceport
        \Parameters -Name HwTimeout -Value 0x00007530 -Verbose
    }
}
```

4. Reboot all nodes to apply all of the changes.
Step 3: Create Failover Cluster

S2D is a feature that is enabled on an existing failover cluster. After you enable S2D on a failover cluster, it takes control of the local storage of each node in the cluster. For this reason, we recommend that you install a cluster with no storage at creation time, and then enable S2D.

When you create a cluster on AWS, you must assign static IP addresses from each subnet from which a node is deployed. From the console, they must be set as secondary private IP addresses on each node. For this tutorial, we configured 172.16.1.200 and 172.16.3.200 upon deployment of each node.

You can verify and review the cluster configuration with the built-in Test-Cluster command.

Test and verify your cluster configuration

1. Run the Test-Cluster command with the Storage Spaces Direct, Inventory, Network, and System Configuration tests.

```powershell
$report = Test-Cluster -Node $nodes -Include 'Storage Spaces Direct', 'Inventory', 'Network', 'System Configuration'
```

2. Review the test results.

```powershell
$reportFilePath = $report.FullName
Start-Process $reportFilePath
```

3. Create the cluster using New-Cluster. Virtual IPs must be assigned a secondary private IP address from the AWS Console to each respective node.

```powershell
$vips = "172.16.1.200", "172.16.3.200"
New-Cluster -Name S2D -Node $nodes -StaticAddress $vips -NoStorage
```

4. Configure a file share witness.

```powershell
New-Item -ItemType Directory -Path c:\Share\Witness
[string]$DomainName = (Get-WmiObject win32_computersystem).domain
New-SmbShare -Name fsw -Path c:\Share\Witness -FullAccess ($DomainName + "\Domain Computers")
Set-ClusterQuorum -Cluster S2D -FileShareWitness \$env:COMPUTERNAME\fsw
```

Step 4: Enable S2D

When the cluster is ready, enable S2D on one of the nodes using Enable-ClusterS2D as follows. Because we have only one type of disk in our setup (local NVMe), we won't use any disks as a cache disk.

1. Enable S2D on i3 instance types using the Enable-ClusterS2D command.

```powershell
Enable-ClusterS2D -PoolFriendlyName S2DPool -Confirm:$false -SkipEligibilityChecks:$true -CimSession $nodes[0]
```

2. If you are using m5d or r5d instance types with NVMe and EBS, use NVMe disks as cache disks. The command would look like this:
Step 5: Provision Storage

To provision storage, create a storage pool and then create volumes in that pool. To keep thing simple, by default, the `Enable-ClusterS2D` command creates a pool using all of the disks available in the cluster. With this command we configured the storage pool name as "S2D Pool."

After volumes are created, they become accessible to every node in the cluster. The volumes can then be assigned to a specific role in the cluster, such as a file server role; or, they can be assigned as cluster shared volumes (CSV). A CSV is accessible to the entire cluster, which means that every node in this cluster can write-read to this volume.

To improve performance, we recommend you use fixed provisioning and a ReFS file system for CSV. Sector size depends on what type of workloads will be deployed on the cluster. For more information on sector size, see [Cluster Size Recommendations for ReFS and NTFS](#). For improved local read performance, we recommend that you align the CSV with the node hosting your application or workload. You can have multiple CSV and multiple applications spread across nodes.

Create a cluster shared volume (CSV)

- Use the `New-Volume` command to create a new 1TB CSV.

```powershell
$Params = @{
    FriendlyName = 'CSV1';
    FileSystem   = 'CSVFS_ReFS';
    StoragePoolFriendlyName = 'S2DPool';
    Size         = 1TB;
    AllocationUnitSize = 65536;
    ProvisioningType = 'Fixed';
    CimSession   = $nodes[0];
}
New-Volume @Params
```

Step 6: Review the S2D Resources

The S2D resources that you configured are displayed in the Failover Cluster Manager.

To view your CSV

1. Open Server Manager.
2. Choose **Tools, Failover Cluster Manager**.
3. Expand the name of the cluster, expand **Storage**, and choose **Disks**.

   The friendly name, capacity, node hosting the CSV, and other data will be listed. For more information on managing CSVs, see [Use Cluster Shared Volumes in a Failover Cluster](#).

To synthesize a load on your CSV

Use a tool such as **Diskspd Utility**. Connect to one of the cluster nodes with RDP and run the following with the Diskspd tool.

```
$mycsv = (gci C:\ClusterStorage\ | select -First 1).FullName
```
To view the S2D storage performance of the cluster

Use the `Get-StorageHealthReport` command to view the cluster performance on one of the cluster nodes.

1. Open a new PowerShell windows and start your synthesized workload.
2. In your original PowerShell windows, run `Get-StorageSubSystem *cluster* | Get-StorageHealthReport` to see the performance results of the storage subsystem while the workload is running.

```powershell
PS C:\> Get-StorageSubSystem *cluster* | Get-StorageHealthReport
```

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUUsageAverage</td>
<td>60.44 %</td>
</tr>
<tr>
<td>CapacityPhysicalPooledAvailable</td>
<td>9.82 GB</td>
</tr>
<tr>
<td>CapacityPhysicalPooledTotal</td>
<td>13.82 TB</td>
</tr>
<tr>
<td>CapacityPhysicalTotal</td>
<td>13.82 TB</td>
</tr>
<tr>
<td>CapacityPhysicalUnpooled</td>
<td>0 B</td>
</tr>
<tr>
<td>CapacityVolumesAvailable</td>
<td>1.89 TB</td>
</tr>
<tr>
<td>CapacityVolumesTotal</td>
<td>2 TB</td>
</tr>
<tr>
<td>IOLatencyAverage</td>
<td>257.56 ms</td>
</tr>
<tr>
<td>IOLatencyRead</td>
<td>255.87 ms</td>
</tr>
<tr>
<td>IOLatencyWrite</td>
<td>259.25 ms</td>
</tr>
<tr>
<td>IOPSTotal</td>
<td>128582.85 /S</td>
</tr>
<tr>
<td>IOPSSWrite</td>
<td>64255.49 /S</td>
</tr>
<tr>
<td>IOTThroughputRead</td>
<td>251.28 MB/S</td>
</tr>
<tr>
<td>IOTThroughputTotal</td>
<td>502.28 MB/S</td>
</tr>
<tr>
<td>IOTThroughputWrite</td>
<td>251 MB/S</td>
</tr>
<tr>
<td>MemoryAvailable</td>
<td>477.77 GB</td>
</tr>
<tr>
<td>MemoryTotal</td>
<td>488 GB</td>
</tr>
</tbody>
</table>

**Step 7: Clean Up**

If you followed the tutorial to create a highly available storage cluster using S2D in EC2 Windows, you created a Storage Spaces Direct cluster of two instances from a bastion server, which also serves as a file share witness to the cluster. You are charged for each hour or partial hour that you keep your instances running. When you no longer need your cluster, use the EC2 Console or the AWS Tools for Windows to delete the resources you created for this project. Do this by deleting the cluster from the failover cluster management mmc, terminating the instances, and deleting the computer objects for the cluster and its respective nodes from your Active Directory.
Additional Resources

Storage Spaces Direct Calculator (Preview)
Planning Storage Spaces Direct
Storage Spaces Direct Overview
Fault Tolerance and Storage Efficiency in Storage Spaces Direct
Resources and Tags

Amazon EC2 provides different resources that you can create and use. Some of these resources include images, instances, volumes, and snapshots. When you create a resource, we assign the resource a unique resource ID.

Some resources can be tagged with values that you define, to help you organize and identify them.

The following topics describe resources and tags, and how you can work with them.

Contents

- Resource Locations (p. 955)
- Resource IDs (p. 956)
- Listing and Filtering Your Resources (p. 961)
- Tagging Your Amazon EC2 Resources (p. 964)
- Amazon EC2 Service Limits (p. 974)
- Amazon EC2 Usage Reports (p. 976)

Resource Locations

Some resources can be used in all regions (global), and some resources are specific to the region or Availability Zone in which they reside.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS account</td>
<td>Global</td>
<td>You can use the same AWS account in all regions.</td>
</tr>
<tr>
<td>Key pairs</td>
<td>Global or Regional</td>
<td>The key pairs that you create using Amazon EC2 are tied to the region where you created them. You can create your own RSA key pair and upload it to the region in which you want to use it; therefore, you can make your key pair globally available by uploading it to each region. For more information, see Amazon EC2 Key Pairs and Windows Instances (p. 578).</td>
</tr>
<tr>
<td>Amazon EC2 resource identifiers</td>
<td>Regional</td>
<td>Each resource identifier, such as an AMI ID, instance ID, EBS volume ID, or EBS snapshot ID, is tied to its region and can be used only in the region where you created the resource.</td>
</tr>
<tr>
<td>User-supplied resource names</td>
<td>Regional</td>
<td>Each resource name, such as a security group name or key pair name, is tied to its region and can be used only in the region where you created the resource. Although you can create resources with the same name in multiple regions, they aren’t related to each other.</td>
</tr>
<tr>
<td>AMIs</td>
<td>Regional</td>
<td>An AMI is tied to the region where its files are located within Amazon S3. You can copy an AMI from one region to another. For more information, see Copying an AMI (p. 72).</td>
</tr>
</tbody>
</table>
Resource IDs

When resources are created, we assign each resource a unique resource ID. You can use resource IDs to find your resources in the Amazon EC2 console. If you are using a command line tool or the Amazon EC2 API to work with Amazon EC2, resource IDs are required for certain commands. For example, if you are using the stop-instances AWS CLI command to stop an instance, you must specify the instance ID in the command.

Resource ID Length

A resource ID takes the form of a resource identifier (such as snap for a snapshot) followed by a hyphen and a unique combination of letters and numbers. Starting in January 2016, we’re gradually introducing longer length IDs for Amazon EC2 and Amazon EBS resource types. The length of the alphanumeric character combination was in an 8-character format; the new IDs are in a 17-character format, for example, i-1234567890abcdef0 for an instance ID.

Supported resource types have an opt-in period, during which you can choose a resource ID format, and a deadline date, after which the resource defaults to the longer ID format. After the deadline has passed for a specific resource type, you can no longer disable the longer ID format for that resource type.

Different resource types have different opt-in periods and deadline dates. The following table lists the supported resource types, along with their opt-in periods and deadline dates.

<table>
<thead>
<tr>
<th>Resource type</th>
<th>Opt-in period</th>
<th>Deadline date</th>
</tr>
</thead>
<tbody>
<tr>
<td>instance</td>
<td>snapshot</td>
<td>reservation</td>
</tr>
<tr>
<td>bundle</td>
<td>conversion-task</td>
<td>customer-gateway</td>
</tr>
</tbody>
</table>
# Working with Longer IDs

The Amazon Elastic Compute Cloud (Amazon EC2) User Guide for Windows Instances includes information about working with longer IDs.

## Resource Types

<table>
<thead>
<tr>
<th>Resource type</th>
<th>Opt-in period</th>
<th>Deadline date</th>
</tr>
</thead>
<tbody>
<tr>
<td>elastic-ip-association</td>
<td></td>
<td></td>
</tr>
<tr>
<td>export-task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>flow-log</td>
<td></td>
<td></td>
</tr>
<tr>
<td>image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>import-task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>internet-gateway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>network-acl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>network-acl-association</td>
<td></td>
<td></td>
</tr>
<tr>
<td>network-interface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>network-interface-attachment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prefix-list</td>
<td></td>
<td></td>
</tr>
<tr>
<td>route-table</td>
<td></td>
<td></td>
</tr>
<tr>
<td>route-table-association</td>
<td></td>
<td></td>
</tr>
<tr>
<td>security-group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>subnet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>subnet-cidr-block-association</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vpc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vpc-cidr-block-association</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vpc-endpoint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vpc-peering-connection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vpn-connection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vpn-gateway</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## During the Opt-in Period

You can enable or disable longer IDs for a resource at any time during the opt-in period. After you've enabled longer IDs for a resource type, any new resources that you create are created with a longer ID.

**Note**

A resource ID does not change after it's created. Therefore, enabling or disabling longer IDs during the opt-in period does not affect your existing resource IDs.

Depending on when you created your AWS account, supported resource types may default to using longer IDs. However, you can opt out of using longer IDs until the deadline date for that resource type. For more information, see Longer EC2 and EBS Resource IDs in the Amazon EC2 FAQs.

## After the Deadline Date

You can't disable longer IDs for a resource type after its deadline date has passed. Any new resources that you create are created with a longer ID.

## Working with Longer IDs

You can enable or disable longer IDs per IAM user and IAM role. By default, an IAM user or role defaults to the same settings as the root user.

### Topics

- Viewing Longer ID Settings (p. 957)
- Modifying Longer ID Settings (p. 958)

## Viewing Longer ID Settings

You can use the console and command line tools to view the resource types that support longer IDs.

### To view your longer ID settings using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation bar at the top of the screen, select the region for which to view your longer ID settings.
3. From the dashboard, under **Account Attributes**, choose **Resource ID length management**.
4. Expand **Advanced Resource ID Management** to view the resource types that support longer IDs and their deadline dates.

**To view your longer ID settings using the command line**

Use one of the following commands:

- **describe-id-format** (AWS CLI)

  ```bash
  aws ec2 describe-id-format --region region
  ```

- **Get-EC2IdFormat** (AWS Tools for Windows PowerShell)

  ```powershell
  Get-EC2IdFormat -Region region
  ```

**To view longer ID settings for a specific IAM user or IAM role using the command line**

Use one of the following commands and specify the ARN of an IAM user, IAM role, or root account user in the request.

- **describe-identity-id-format** (AWS CLI)

  ```bash
  aws ec2 describe-identity-id-format --principal-arn arn-of-iam-principal --region region
  ```

- **Get-EC2IdentityIdFormat** (AWS Tools for Windows PowerShell)

  ```powershell
  Get-EC2IdentityIdFormat -PrincipalArn arn-of-iam-principal -Region region
  ```

**To view the aggregated longer ID settings for a specific region using the command line**

Use the **describe-aggregate-id-format** AWS CLI command to view the aggregated longer ID setting for the entire region, as well as the aggregated longer ID setting of all ARNs for each resource type. This command is useful for performing a quick audit to determine whether a specific region is fully opted in for longer IDs.

```bash
aws ec2 describe-aggregate-id-format --region region
```

**To identify users who have explicitly defined custom longer ID settings**

Use the **describe-principal-id-format** AWS CLI command to view the longer ID format settings for the root user and all IAM roles and IAM users that have explicitly specified a longer ID preference. This command is useful for identifying IAM users and IAM roles that have overridden the default longer ID settings.

```bash
aws ec2 describe-principal-id-format --region region
```

**Modifying Longer ID Settings**

You can use the console and command line tools to modify longer ID settings for resource types that are still within their opt-in period.

**Note**

The AWS CLI and AWS Tools for Windows PowerShell commands in this section are per-region only. They apply to the default region unless otherwise specified. To modify the settings for other regions, include the `region` parameter in the command.
To modify longer ID settings using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation bar at the top of the screen, select the region for which to modify the longer ID settings.
3. From the dashboard, under Account Attributes, choose Resource ID length management.
4. Do one of the following:
   - To enable longer IDs for all supported resource types for all IAM users across all regions, choose Switch to longer IDs, Yes, switch to longer IDs.  
     
     Important
     IAM users and IAM roles need the ec2:ModifyIdentityIdFormat permission to perform this action.
   - To modify longer ID settings for a specific resource type for your IAM user account, expand Advanced Resource ID Management, and then select the corresponding check box in the My IAM Role/User column to enable longer IDs, or clear the check box to disable longer IDs.
   - To modify longer ID settings for a specific resource type for all IAM users, expand Advanced Resource ID Management, and then select the corresponding check box in the All IAM Roles/Users column to enable longer IDs, or clear the check box to disable longer IDs.

To modify longer ID settings for your IAM user account using the command line

Use one of the following commands:

   • modify-id-format (AWS CLI)
     ```
     aws ec2 modify-id-format --resource resource_type --use-long-ids
     ```
     You can also use the command to modify the longer ID settings for all supported resource types. To do this, replace the resource_type parameter with all-current.
     ```
     aws ec2 modify-id-format --resource all-current --use-long-ids
     ```

   Note
   If you're using these commands as the root user, then changes apply to the entire AWS account, unless an IAM user or role explicitly overrides these settings for themselves.

   • Edit-EC2IdFormat (AWS Tools for Windows PowerShell)
     ```
     Edit-EC2IdFormat -Resource resource_type -UseLongId boolean
     ```
     You can also use the command to modify the longer ID settings for all supported resource types. To do this, replace the resource_type parameter with all-current.
     ```
     Edit-EC2IdFormat -Resource all-current -UseLongId boolean
     ```

To modify longer ID settings for a specific IAM user or IAM role using the command line

Use one of the following commands and specify the ARN of an IAM user, IAM role, or root user in the request.
Controlling Access to Longer ID Settings

By default, IAM users and roles do not have permission to use the following actions unless they're explicitly granted permission through their associated IAM policies:

- ec2:DescribeIdFormat
- ec2:DescribeIdentityIdFormat
- ec2:DescribeAggregateIdFormat
- ec2:DescribePrincipalIdFormat
- ec2:ModifyIdFormat
- ec2:ModifyIdentityIdFormat

For example, an IAM role may have permission to use all Amazon EC2 actions through an "Action": "ec2:*" element in the policy statement.

To prevent IAM users and roles from viewing or modifying the longer resource ID settings for themselves or other users and roles in your account, ensure that the IAM policy contains the following statement:

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "ec2:ModifyIdFormat",
        "ec2:DescribeIdFormat",
        "ec2:ModifyIdentityIdFormat",
        "ec2:DescribeIdentityIdFormat"
      ],
      "Resource": "arn-of-iam-principal"
    }
  ]
}
```
We do not support resource-level permissions for the following actions:

- `ec2:DescribeIdFormat`
- `ec2:DescribeIdentityIdFormat`
- `ec2:DescribeAggregateIdFormat`
- `ec2:DescribePrincipalIdFormat`
- `ec2:ModifyIdFormat`
- `ec2:ModifyIdentityIdFormat`

### Listing and Filtering Your Resources

You can get a list of some types of resource using the Amazon EC2 console. You can get a list of each type of resource using its corresponding command or API action. If you have many resources, you can filter the results to include only the resources that match certain criteria.

#### Contents

- Advanced Search (p. 961)
- Listing Resources Using the Console (p. 962)
- Filtering Resources Using the Console (p. 963)
- Listing and Filtering Using the CLI and API (p. 964)

### Advanced Search

Advanced search allows you to search using a combination of filters to achieve precise results. You can filter by keywords, user-defined tag keys, and predefined resource attributes.

The specific search types available are:

- **Search by keyword**

  To search by keyword, type or paste what you’re looking for in the search box, and then choose Enter. For example, to search for a specific instance, you can type the instance ID.

- **Search by fields**

  You can also search by fields, tags, and attributes associated with a resource. For example, to find all instances in the stopped state:
  1. In the search box, start typing **Instance State**. As you type, you’ll see a list of suggested fields.
  2. Select **Instance State** from the list.
  3. Select **Stopped** from the list of suggested values.
  4. To further refine your list, select the search box for more search options.

- **Advanced search**
You can create advanced queries by adding multiple filters. For example, you can search by tags and see instances for the Flying Mountain project running in the Production stack, and then search by attributes to see all t2.micro instances, or all instances in us-west-2a, or both.

- **Inverse search**

You can search for resources that do not match a specified value. For example, to list all instances that are not terminated, search by the **Instance State** field, and prefix the Terminated value with an exclamation mark (!).

- **Partial search**

When searching by field, you can also enter a partial string to find all resources that contain the string in that field. For example, search by **Instance Type**, and then type `t2` to find all t2.micro, t2.small or t2.medium instances.

- **Regular expression**

  Regular expressions are useful when you need to match the values in a field with a specific pattern. For example, search by the Name tag, and then type `^s.*` to see all instances with a Name tag that starts with an 's'. Regular expression search is not case-sensitive.

After you have the precise results of your search, you can bookmark the URL for easy reference. In situations where you have thousands of instances, filters and bookmarks can save you a great deal of time; you don’t have to run searches repeatedly.

**Combining search filters**

In general, multiple filters with the same key field (for example, tag:Name, search, Instance State) are automatically joined with OR. This is intentional, as the vast majority of filters would not be logical if they were joined with AND. For example, you would get zero results for a search on Instance State=running AND Instance State=stopped. In many cases, you can granulate the results by using complementary search terms on different key fields, where the AND rule is automatically applied instead. If you search for tag: Name:=All values and tag:Instance State=running, you get search results that contain both those criteria. To fine-tune your results, simply remove one filter in the string until the results fit your requirements.

**Listing Resources Using the Console**

You can view the most common Amazon EC2 resource types using the console. To view additional resources, use the command line interface or the API actions.

**To list EC2 resources using the console**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. In the navigation pane, choose the option that corresponds to the resource, such as **AMIs** or **Instances**.
Filtering Resources Using the Console

You can perform filtering and sorting of the most common resource types using the Amazon EC2 console. For example, you can use the search bar on the instances page to sort instances by tags, attributes, or keywords.

You can also use the search field on each page to find resources with specific attributes or values. You can use regular expressions to search on partial or multiple strings. For example, to find all instances that are using the MySG security group, enter MySG in the search field. The results will include any values that contain MySG as a part of the string, such as MySG2 and MySG3. To limit your results to MySG only, enter \bMySG\b in the search field. To list all the instances whose type is either m1.small or m1.large, enter m1.small|m1.large in the search field.

To list volumes in the us-east-1b Availability Zone with a status of available

1. In the navigation pane, choose Volumes.
2. Click on the search box, select Attachment Status from the menu, and then select Detached. (A detached volume is available to be attached to an instance in the same Availability Zone.)
3. Click on the search box again, select State, and then select Available.
4. Click on the search box again, select Availability Zone, and then select us-east-1b.
5. Any volumes that meet this criteria are displayed.
To list public 64-bit Windows AMIs backed by Amazon EBS

1. In the navigation pane, choose AMIs.
2. In the Filter pane, select Public images, EBS images, and then Windows from the Filter lists.
3. Type x86_64 in the search field.
4. Any AMIs that meet this criteria are displayed.

Listing and Filtering Using the CLI and API

Each resource type has a corresponding CLI command or API request that you use to list resources of that type. For example, you can list Amazon Machine Images (AMIs) using `ec2-describe-images` or `DescribeImages`. The response contains information for all your resources.

The resulting lists of resources can be long, so you might want to filter the results to include only the resources that match certain criteria. You can specify multiple filter values, and you can also specify multiple filters. For example, you can list all the instances whose type is either `m1.small` or `m1.large`, and that have an attached EBS volume that is set to delete when the instance terminates. The instance must match all your filters to be included in the results.

You can also use wildcards with the filter values. An asterisk (*) matches zero or more characters, and a question mark (?) matches zero or one character.

For example, you can use `database` as the filter value to get only the EBS snapshots whose description equals `database`. If you specify `*database*`, then all snapshots whose description includes `database` are returned. If you specify `database?`, then only the snapshots whose description matches one of the following patterns are returned: equals `database` or equals `database` followed by one character.

The number of question marks determines the maximum number of characters to include in results. For example, if you specify `database????`, then only the snapshots whose description equals `database` followed by up to four characters are returned. Descriptions with five or more characters following `database` are excluded from the search results.

Filter values are case sensitive. We support only exact string matching, or substring matching (with wildcards). If a resulting list of resources is long, using an exact string filter may return the response faster.

Your search can include the literal values of the wildcard characters; you just need to escape them with a backslash before the character. For example, a value of `\*amazon\?\\` searches for the literal string `*amazon?\`.

For a list of supported filters per Amazon EC2 resource, see the relevant documentation:

- For the AWS CLI, see the relevant `describe` command in the [AWS CLI Command Reference](https://docs.aws.amazon.com/cli/latest/reference/ec2/describe-images.html).
- For Windows PowerShell, see the relevant `Get` command in the [AWS Tools for PowerShell Cmdlet Reference](https://docs.aws.amazon.com/powershell/latest/reference/cmdlets/powershell/amazon/ec2/).  
- For the Query API, see the relevant `Describe` API action in the [Amazon EC2 API Reference](https://docs.aws.amazon.com/AmazonEC2/latest/APIReference/).  

Tagging Your Amazon EC2 Resources

To help you manage your instances, images, and other Amazon EC2 resources, you can optionally assign your own metadata to each resource in the form of **tags**. This topic describes tags and shows you how to create them.
Warning
Tag keys and their values are returned by many different API calls. Denying access to DescribeTags doesn't automatically deny access to tags returned by other APIs. As a best practice, we recommend that you do not include sensitive data in your tags.

Contents
- Tag Basics (p. 965)
- Tagging Your Resources (p. 966)
- Tag Restrictions (p. 968)
- Tagging Your Resources for Billing (p. 969)
- Working with Tags Using the Console (p. 969)
- Working with Tags Using the CLI or API (p. 972)

Tag Basics

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value, both of which you define.

Tags enable you to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. This is useful when you have many resources of the same type—you can quickly identify a specific resource based on the tags you've assigned to it. For example, you could define a set of tags for your account's Amazon EC2 instances that helps you track each instance's owner and stack level.

The following diagram illustrates how tagging works. In this example, you've assigned two tags to each of your instances—one tag with the key Owner and another with the key Stack. Each tag also has an associated value.
We recommend that you devise a set of tag keys that meets your needs for each resource type. Using a consistent set of tag keys makes it easier for you to manage your resources. You can search and filter the resources based on the tags you add.

Tags don't have any semantic meaning to Amazon EC2 and are interpreted strictly as a string of characters. Also, tags are not automatically assigned to your resources. You can edit tag keys and values, and you can remove tags from a resource at any time. You can set the value of a tag to an empty string, but you can't set the value of a tag to null. If you add a tag that has the same key as an existing tag on that resource, the new value overwrites the old value. If you delete a resource, any tags for the resource are also deleted.

You can work with tags using the AWS Management Console, the AWS CLI, and the Amazon EC2 API.

If you're using AWS Identity and Access Management (IAM), you can control which users in your AWS account have permission to create, edit, or delete tags. For more information, see Controlling Access to Amazon EC2 Resources (p. 597).

Tagging Your Resources

You can tag most Amazon EC2 resources that already exist in your account. The table (p. 966) below lists the resources that support tagging.

If you're using the Amazon EC2 console, you can apply tags to resources by using the Tags tab on the relevant resource screen, or you can use the Tags screen. Some resource screens enable you to specify tags for a resource when you create the resource; for example, a tag with a key of Name and a value that you specify. In most cases, the console applies the tags immediately after the resource is created (rather than during resource creation). The console may organize resources according to the Name tag, but this tag doesn't have any semantic meaning to the Amazon EC2 service.

If you're using the Amazon EC2 API, the AWS CLI, or an AWS SDK, you can use the CreateTags EC2 API action to apply tags to existing resources. Additionally, some resource-creating actions enable you to specify tags for a resource when the resource is created. If tags cannot be applied during resource creation, we roll back the resource creation process. This ensures that resources are either created with tags or not created at all, and that no resources are left untagged at any time. By tagging resources at the time of creation, you can eliminate the need to run custom tagging scripts after resource creation.

The following table describes the Amazon EC2 resources that can be tagged, and the resources that can be tagged on creation using the Amazon EC2 API, the AWS CLI, or an AWS SDK.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Supports tags</th>
<th>Supports tagging on creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFI</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>AMI</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Bundle task</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Capacity Reservation</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Client VPN endpoint</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Client VPN route</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Customer gateway</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Dedicated Host</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DHCP option</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Resource</td>
<td>Supports tags</td>
<td>Supports tagging on creation</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>EBS snapshot</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>EBS volume</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>EC2 Fleet</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Egress-only internet gateway</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Elastic IP address</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Dedicated Host Reservation</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Instance</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Instance store volume</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Internet gateway</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Key pair</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Launch template</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Launch template version</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>NAT gateway</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Network ACL</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Network interface</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Placement group</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Reserved Instance</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Reserved Instance listing</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Route table</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Spot Instance request</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Security group</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Subnet</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Traffic Mirror filter</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Traffic Mirror session</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Traffic Mirror target</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transit gateway</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transit gateway route table</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transit gateway VPC attachment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Virtual private gateway</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>VPC</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>VPC endpoint</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
### Tag Restrictions

<table>
<thead>
<tr>
<th>Resource</th>
<th>Supports tags</th>
<th>Supports tagging on creation</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPC endpoint service</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>VPC endpoint service configuration</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>VPC flow log</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>VPC peering connection</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>VPN connection</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

You can tag instances and volumes on creation using the Amazon EC2 Launch Instances wizard in the Amazon EC2 console. You can tag your EBS volumes on creation using the Volumes screen, or EBS snapshots using the Snapshots screen. Alternatively, use the resource-creating Amazon EC2 APIs (for example, `RunInstances`) to apply tags when creating your resource.

You can apply tag-based resource-level permissions in your IAM policies to the Amazon EC2 API actions that support tagging on creation to implement granular control over the users and groups that can tag resources on creation. Your resources are properly secured from creation—tags are applied immediately to your resources, therefore any tag-based resource-level permissions controlling the use of resources are immediately effective. Your resources can be tracked and reported on more accurately. You can enforce the use of tagging on new resources, and control which tag keys and values are set on your resources.

You can also apply resource-level permissions to the `CreateTags` and `DeleteTags` Amazon EC2 API actions in your IAM policies to control which tag keys and values are set on your existing resources. For more information, see Supported Resource-Level Permissions for Amazon EC2 API Actions (p. 608) and Example Policies for Working with the AWS CLI or an AWS SDK (p. 611).

For more information about tagging your resources for billing, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

### Tag Restrictions

The following basic restrictions apply to tags:

- Maximum number of tags per resource – 50
- For each resource, each tag key must be unique, and each tag key can have only one value.
- Maximum key length – 128 Unicode characters in UTF-8
- Maximum value length – 256 Unicode characters in UTF-8
- Although EC2 allows for any character in its tags, other services are more restrictive. The allowed characters across services are: letters, numbers, and spaces representable in UTF-8, and the following characters: + - = . _ : / @.
- Tag keys and values are case-sensitive.
- The `aws:` prefix is reserved for AWS use. If a tag has a tag key with this prefix, then you can't edit or delete the tag's key or value. Tags with the `aws:` prefix do not count against your tags per resource limit.

You can't terminate, stop, or delete a resource based solely on its tags; you must specify the resource identifier. For example, to delete snapshots that you tagged with a tag key called `DeleteMe`, you must use the `DeleteSnapshots` action with the resource identifiers of the snapshots, such as `snap-1234567890abcdef0`.

You can tag public or shared resources, but the tags you assign are available only to your AWS account and not to the other accounts sharing the resource.
You can't tag all resources. For more information, see Tagging Support for Amazon EC2 Resources (p. 966).

Tagging Your Resources for Billing

You can use tags to organize your AWS bill to reflect your own cost structure. To do this, sign up to get your AWS account bill with tag key values included. For more information about setting up a cost allocation report with tags, see The Monthly Cost Allocation Report in AWS Billing and Cost Management User Guide. To see the cost of your combined resources, you can organize your billing information based on resources that have the same tag key values. For example, you can tag several resources with a specific application name, and then organize your billing information to see the total cost of that application across several services. For more information, see Using Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

Note
If you've just enabled reporting, data for the current month is available for viewing after 24 hours.

Cost allocation tags can indicate which resources are contributing to costs, but deleting or deactivating resources doesn't always reduce costs. For example, snapshot data that is referenced by another snapshot is preserved, even if the snapshot that contains the original data is deleted. For more information, see Amazon Elastic Block Store Volumes and Snapshots in the AWS Billing and Cost Management User Guide.

Note
Elastic IP addresses that are tagged do not appear on your cost allocation report.

Working with Tags Using the Console

Using the Amazon EC2 console, you can see which tags are in use across all of your Amazon EC2 resources in the same Region. You can view tags by resource and by resource type, and you can also view how many items of each resource type are associated with a specified tag. You can also use the Amazon EC2 console to apply or remove tags from one or more resources at a time.

For more information about using filters when listing your resources, see Listing and Filtering Your Resources (p. 961).

For ease of use and best results, use Tag Editor in the AWS Management Console, which provides a central, unified way to create and manage your tags. For more information, see Working with Tag Editor in Getting Started with the AWS Management Console.

Contents
- Displaying Tags (p. 969)
- Adding and Deleting Tags on an Individual Resource (p. 970)
- Adding and Deleting Tags to a Group of Resources (p. 971)
- Adding a Tag When You Launch an Instance (p. 971)
- Filtering a List of Resources by Tag (p. 972)

Displaying Tags

You can display tags in two different ways in the Amazon EC2 console. You can display the tags for an individual resource or for all resources.

Displaying Tags for Individual Resources

When you select a resource-specific page in the Amazon EC2 console, it displays a list of those resources. For example, if you select Instances from the navigation pane, the console displays a list of Amazon EC2...
instances. When you select a resource from one of these lists (for example, an instance), if the resource supports tags, you can view and manage its tags. On most resource pages, you can view the tags in the **Tags** tab on the details pane.

You can add a column to the resource list that displays all values for tags with the same key. This column enables you to sort and filter the resource list by the tag. There are two ways to add a new column to the resource list to display your tags.

- On the **Tags** tab, select **Show Column**. A new column is added to the console.
- Choose the **Show/Hide Columns** gear-shaped icon, and in the **Show/Hide Columns** dialog box, select the tag key under **Your Tag Keys**.

### Displaying Tags for All Resources

You can display tags across all resources by selecting **Tags** from the navigation pane in the Amazon EC2 console. The following image shows the **Tags** pane, which lists all tags in use by resource type.

![Tags pane](image)

### Adding and Deleting Tags on an Individual Resource

You can manage tags for an individual resource directly from the resource's page.

**To add a tag to an individual resource**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. From the navigation bar, select the Region that meets your needs. This choice is important because some Amazon EC2 resources can be shared between Regions, while others can't. For more information, see [Resource Locations (p. 955)](#).
3. In the navigation pane, select a resource type (for example, **Instances**).
4. Select the resource from the resource list and choose **Tags**, **Add/Edit Tags**.
5. In the **Add/Edit Tags** dialog box, specify the key and value for each tag, and then choose **Save**.

**To delete a tag from an individual resource**

1. Open the Amazon EC2 console at [https://console.aws.amazon.com/ec2/](https://console.aws.amazon.com/ec2/).
2. From the navigation bar, select the Region that meets your needs. This choice is important because some Amazon EC2 resources can be shared between Regions, while others can't. For more information, see [Resource Locations (p. 955)](#).
3. In the navigation pane, choose a resource type (for example, Instances).
4. Select the resource from the resource list and choose Tags.
5. Choose Add/Edit Tags, select the Delete icon for the tag, and choose Save.

Adding and Deleting Tags to a Group of Resources

To add a tag to a group of resources

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region that meets your needs. This choice is important because some Amazon EC2 resources can be shared between Regions, while others can't. For more information, see Resource Locations (p. 955).
3. In the navigation pane, choose Tags.
4. At the top of the content pane, choose Manage Tags.
5. For Filter, select the type of resource (for example, instances) to which to add tags.
6. In the resources list, select the check box next to each resource to which to add tags.
7. Under Add Tag, for Key and Value, type the tag key and values, and then choose Add Tag.

Note
If you add a new tag with the same tag key as an existing tag, the new tag overwrites the existing tag.

To remove a tag from a group of resources

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select the Region that meets your needs. This choice is important because some Amazon EC2 resources can be shared between Regions, while others can't. For more information, see Resource Locations (p. 955).
3. In the navigation pane, choose Tags, Manage Tags.
4. To view the tags in use, select the Show/Hide Columns gear-shaped icon, and in the Show/Hide Columns dialog box, select the tag keys to view and choose Close.
5. For Filter, select the type of resource (for example, instances) from which to remove tags.
6. In the resource list, select the check box next to each resource from which to remove tags.
7. Under Remove Tag, for Key, type the tag's name and choose Remove Tag.

Adding a Tag When YouLaunch an Instance

To add a tag using the Launch Wizard

1. From the navigation bar, select the Region for the instance. This choice is important because some Amazon EC2 resources can be shared between Regions, while others can't. Select the Region that meets your needs. For more information, see Resource Locations (p. 955).
2. Choose Launch Instance.
3. The Choose an Amazon Machine Image (AMI) page displays a list of basic configurations called Amazon Machine Images (AMIs). Select the AMI to use and choose Select. For more information about selecting an AMI, see Finding a Windows AMI (p. 51).
4. On the Configure Instance Details page, configure the instance settings as necessary, and then choose Next: Add Storage.
5. On the **Add Storage** page, you can specify additional storage volumes for your instance. Choose **Next: Add Tags** when done.

6. On the **Add Tags** page, specify tags for the instance, the volumes, or both. Choose **Add another tag** to add more than one tag to your instance. Choose **Next: Configure Security Group** when you are done.

7. On the **Configure Security Group** page, you can choose from an existing security group that you own, or let the wizard create a new security group for you. Choose **Review and Launch** when you are done.

8. Review your settings. When you're satisfied with your selections, choose **Launch**. Select an existing key pair or create a new one, select the acknowledgment check box, and then choose **Launch Instances**.

### Filtering a List of Resources by Tag

You can filter your list of resources based on one or more tag keys and tag values.

**To filter a list of resources by tag**

1. Display a column for the tag as follows:
   
   a. Select a resource.
   
   b. In the details pane, choose **Tags**.
   
   c. Locate the tag in the list and choose **Show Column**.

2. Choose the filter icon in the top right corner of the column for the tag to display the filter list.

3. Select the tag values, and then choose **Apply Filter** to filter the results list.

**Note**

For more information about filters, see [Listing and Filtering Your Resources](p. 961).

### Working with Tags Using the CLI or API

Use the following to add, update, list, and delete the tags for your resources. The corresponding documentation provides examples.

<table>
<thead>
<tr>
<th>Task</th>
<th>AWS CLI</th>
<th>AWS Tools for Windows PowerShell</th>
<th>API Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add or overwrite one or more tags.</td>
<td>create-tags</td>
<td>New-EC2Tag</td>
<td>CreateTags</td>
</tr>
<tr>
<td>Delete one or more tags.</td>
<td>delete-tags</td>
<td>Remove-EC2Tag</td>
<td>DeleteTags</td>
</tr>
<tr>
<td>Describe one or more tags.</td>
<td>describe-tags</td>
<td>Get-EC2Tag</td>
<td>DescribeTags</td>
</tr>
</tbody>
</table>

You can also filter a list of resources according to their tags. The following examples demonstrate how to filter your instances using tags with the **describe-instances** command.

**Note**

The way you enter JSON-formatted parameters on the command line differs depending on your operating system. Linux, macOS, or Unix and Windows PowerShell use the single quote (') to enclose the JSON data structure. Omit the single quotes when using the commands with the Windows command line. For more information, see [Specifying Parameter Values for the AWS Command Line Interface](p. 972).
Example 1: Describe instances with the specified tag key

The following command describes the instances with a Stack tag, regardless of the value of the tag.

```
aws ec2 describe-instances --filters Name=tag-key,Values=Stack
```

Example 2: Describe instances with the specified tag

The following command describes the instances with the tag Stack=production.

```
aws ec2 describe-instances --filters Name=tag:Stack,Values=production
```

Example 3: Describe instances with the specified tag value

The following command describes the instances with a tag with the value production, regardless of the tag key.

```
aws ec2 describe-instances --filters Name=tag-value,Values=production
```

Some resource-creating actions enable you to specify tags when you create the resource. The following actions support tagging on creation.

<table>
<thead>
<tr>
<th>Task</th>
<th>AWS CLI</th>
<th>AWS Tools for Windows PowerShell</th>
<th>API Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Launch one or more instances.</td>
<td>run-instances</td>
<td>New-EC2Instance</td>
<td>RunInstances</td>
</tr>
<tr>
<td>Create an EBS volume.</td>
<td>create-volume</td>
<td>New-EC2Volume</td>
<td>CreateVolume</td>
</tr>
</tbody>
</table>

The following examples demonstrate how to apply tags when you create resources.

Example 4: Launch an instance and apply tags to the instance and volume

The following command launches an instance and applies a tag with a key of `webserver` and value of `production` to the instance. The command also applies a tag with a key of `cost-center` and a value of `cc123` to any EBS volume that's created (in this case, the root volume).

```
aws ec2 run-instances --image-id ami-abc12345 --count 1 --instance-type t2.micro --key-name MyKeyPair --subnet-id subnet-6e7f829e --tag-specifications 'ResourceType=instance,Tags=[{Key=webserver,Value=production}]' 'ResourceType=volume,Tags=[{Key=cost-center,Value=cc123}]'
```

You can apply the same tag keys and values to both instances and volumes during launch. The following command launches an instance and applies a tag with a key of `cost-center` and a value of `cc123` to both the instance and any EBS volume that's created.

```
aws ec2 run-instances --image-id ami-abc12345 --count 1 --instance-type t2.micro --key-name MyKeyPair --subnet-id subnet-6e7f829e --tag-specifications 'ResourceType=instance,Tags=[{Key=cost-center,Value=cc123}]' 'ResourceType=volume,Tags=[{Key=cost-center,Value=cc123}]'
```

Example 5: Create a volume and apply a tag
The following command creates a volume and applies two tags: purpose = production, and cost-center = cc123.

```bash
aws ec2 create-volume --availability-zone us-east-1a --volume-type gp2 --size 80 --tag-specifications 'ResourceType=volume,Tags=[{Key=purpose,Value=production},{Key=cost-center,Value=cc123}]
```

**Example 6: Add a tag to a resource**

This example adds the tag Stack=production to the specified image, or overwrites an existing tag for the AMI where the tag key is Stack. If the command succeeds, no output is returned.

```bash
aws ec2 create-tags --resources ami-78a54011 --tags Key=Stack,Value=production
```

**Example 7: Add tags to multiple resources**

This example adds (or overwrites) two tags for an AMI and an instance. One of the tags contains just a key (webserver), with no value (we set the value to an empty string). The other tag consists of a key (stack) and value (Production). If the command succeeds, no output is returned.

```bash
aws ec2 create-tags --resources ami-1a2b3c4d i-1234567890abcdef0 --tags Key=webserver,Value=  Key=stack,Value=Production
```

**Example 8: Add tags with special characters**

This example adds the tag [Group]=test to an instance. The square brackets ([ and ]) are special characters, and must be escaped with a backslash (\). If you are using Windows, surround the value with double quotes ("').

```bash
aws ec2 create-tags --resources i-1234567890abcdef0 --tags Key="[Group]",Value=test
```

If you are using Windows PowerShell, break out the characters with a backslash (\), surround them with double quotes ("'), and then surround the entire key and value structure with single quotes (').

```bash
aws ec2 create-tags --resources i-1234567890abcdef0 --tags 'Key="[Group]",Value=test'
```

## Amazon EC2 Service Limits

Amazon EC2 provides different resources that you can use. These resources include images, instances, volumes, and snapshots. When you create your AWS account, we set default limits (also referred to as quotas) on these resources on a per-Region basis. For example, there is a limit on the number of instances that you can launch in a Region. Therefore, when you launch an instance in the US West (Oregon) Region, the request must not cause your usage to exceed your current instance limit in that Region.

The Amazon EC2 console provides limit information for the resources managed by the Amazon EC2 and Amazon VPC consoles. You can request an increase for many of these limits. Use the limit information that we provide to manage your AWS infrastructure. Plan to request any limit increases in advance of the time that you'll need them.

For more information about the limits for other services, see AWS Service Limits in the Amazon Web Services General Reference.
Viewing Your Current Limits

Use the EC2 Limits page in the Amazon EC2 console to view the current limits for resources provided by Amazon EC2 and Amazon VPC, on a per-Region basis.

To view your current limits

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select a Region.
3. From the navigation pane, choose Limits.
4. Locate the resource in the list. You can use the search fields to filter the list by resource name or resource group. The Current limit column displays the current maximum for the resource for your account.

   Note
   If you’ve opted in to vCPU-based instance limits for On-Demand Instances, you’ll see five vCPU-based On-Demand Instance limits. Each limit specifies the vCPU limit for one or more instance families. For more information, see On-Demand Instance Limits (p. 194).

Requesting a Limit Increase

Use the Limits page in the Amazon EC2 console to request an increase in the limits for resources provided by Amazon EC2 or Amazon VPC, on a per-Region basis.

To request a limit increase

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. From the navigation bar, select a Region.
3. From the navigation pane, choose Limits.
4. Select the resource in the list, and choose Request limit increase.
5. Complete the required fields on the limit increase form. We’ll respond to you using the contact method that you specified.

Limits on Email Sent Using Port 25

Amazon EC2 throttles traffic on port 25 of all instances by default. You can request that this throttle be removed. For more information, see How do I remove the throttle on port 25 from my EC2 instance? in the AWS Knowledge Center.

Amazon EC2 Usage Reports

AWS provides a free reporting tool called Cost Explorer that enables you to analyze the cost and usage of your EC2 instances and the usage of your Reserved Instances.

Cost Explorer is a free tool that you can use to view charts of your usage and costs. You can view data up to the last 13 months, and forecast how much you are likely to spend for the next three months. You can use Cost Explorer to see patterns in how much you spend on AWS resources over time, identify areas that need further inquiry, and see trends that you can use to understand your costs. You also can specify time ranges for the data, and view time data by day or by month.

Here’s an example of some of the questions that you can answer when using Cost Explorer:

- How much am I spending on instances of each instance type?
- How many instance hours are being used by a particular department?
- How is my instance usage distributed across Availability Zones?
- How is my instance usage distributed across AWS accounts?
- How well am I using my Reserved Instances?
- Are my Reserved Instances helping me save money?

To view an Amazon EC2 report in Cost Explorer

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Reports** and select the report to view.

   The report opens in Cost Explorer. It provides a preconfigured view, based on fixed filter settings, that displays information about your usage and cost trends.

   For more information about working with reports in Cost Explorer, including saving reports, see *Analyzing Your Costs with Cost Explorer*. 
AWS Systems Manager for Microsoft System Center VMM

AWS Systems Manager for Microsoft System Center Virtual Machine Manager (SCVMM) provides a simple, easy-to-use interface for managing AWS resources, such as EC2 instances, from Microsoft SCVMM. It is implemented as an add-in for the VMM console. For more information, see AWS Add-ins for Microsoft System Center.

Features

- Administrators can grant permissions to users so that they can manage EC2 instances from SCVMM.
- Users can launch, view, reboot, stop, start, and terminate instances, if they have the required permissions.
- Users can get the passwords for their Windows instances and connect to them using RDP.
- Users can get the public DNS names for their Linux instances and connect to them using SSH.
- Users can import their Hyper-V Windows virtual machines from SCVMM to Amazon EC2.

Limitations

- Users must have an account that they can use to log in to SCVMM.
- You can't import Linux virtual machines from SCVMM to Amazon EC2.
• This is not a comprehensive tool for creating and managing AWS resources. The add-in enables SCVMM users to get started quickly with the basic tasks for managing their EC2 instances. Future releases might support managing additional AWS resources.

## Requirements

- An AWS account
- Microsoft System Center VMM 2012 R2 or System Center VMM 2012 SP1 with the latest update roll-up

## Getting Started

To get started, see the following documentation:

- Setting Up (p. 979)
- Managing EC2 Instances (p. 983)
- Troubleshooting (p. 989)

## Setting Up AWS Systems Manager for Microsoft SCVMM

When you set up AWS Systems Manager, users in your organization can access your AWS resources. The process involves creating accounts, deploying the add-in, and providing your credentials.

### Tasks

- Sign Up for AWS (p. 979)
- Set Up Access for Users (p. 980)
- Deploy the Add-In (p. 982)
- Provide Your AWS Credentials (p. 982)

## Sign Up for AWS

When you sign up for Amazon Web Services, your AWS account is automatically signed up for all services in AWS. You are charged only for the services that you use.

If you have an AWS account already, skip to the next task. If you don’t have an AWS account, use the following procedure to create one.

### To sign up for an AWS account

2. Follow the online instructions.

   Part of the sign-up procedure involves receiving a phone call and entering a verification code on the phone keypad.
Set Up Access for Users

The first time that you use Systems Manager, you must provide AWS credentials. To enable multiple users to access the same AWS account using unique credentials and permissions, create an IAM user for each user. You can create one or more groups with policies that grant permissions to perform limited tasks. Then you can create one or more IAM users, and add each user to the appropriate group.

To create an Administrators group

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Groups and then choose Create New Group.
3. In the Group Name box, specify Administrators and then choose Next Step.
5. Choose Next Step and then choose Create Group.

To create a group with limited access to Amazon EC2

1. Open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Groups and then choose Create New Group.
3. In the Group Name box, specify a meaningful name for the group and then choose Next Step.
4. On the Attach Policy page, do not select an AWS managed policy — choose Next Step, and then choose Create Group.
5. Choose the name of the group you’ve just created. On the Permissions tab, choose Inline Policies, and then click here.
6. Select the Custom Policy radio button and then choose Select.
7. Enter a name for the policy and a policy document that grants limited access to Amazon EC2, and then choose Apply Policy. For example, you can specify one of the following custom policies.

Grant users in this group permission to view information about EC2 instances only

```
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": [
            "ec2:Describe*",
            "iam:ListInstanceProfiles"
         ],
         "Resource": "*"
      }
   ]
}
```

Grant users in this group permission to perform all operations on EC2 instances that are supported by the add-in

```
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": [
            "iam:ListInstanceProfiles", "iam:PassRole",
            "ec2:Describe*", "ec2:CreateKeyPair",
```
Grant users in this group permission to import a VM to Amazon EC2

```json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
      ],
      "Resource": "*"
    }
  ]
}
```

To create an IAM user, get the user's AWS credentials, and grant the user permissions

1. In the navigation pane, choose Users and then choose Add user.
2. Enter a user name.
3. Select the type of access this set of users will have. Select Programmatic access and AWS Management Console access if this user must also access the AWS Management Console.
4. For Console password type, choose one of the following:
   - Autogenerated password. Each user gets a randomly generated password that meets the current password policy in effect (if any). You can view or download the passwords when you get to the Final page.
   - Custom password. Each user is assigned the password that you type in the box.
5. Choose Next: Permissions.
6. On the Set permissions page, choose Add user to group. Select the appropriate group.
7. Choose Next: Review, then Create user.
8. To view the users' access keys (access key IDs and secret access keys), choose Show next to each password and secret access key that you want to see. To save the access keys, choose Download .csv and then save the file to a safe location.
   **Note**
   You cannot retrieve the secret access key after you complete this step; if you misplace it you must create a new one.
9. Choose Close.
Deploy the Add-In

Add-ins for System Center VMM are distributed as .zip files. To deploy the add-in, use the following procedure.

To deploy the add-in

1. From your instance, go to AWS Systems Manager for Microsoft System Center Virtual Machine Manager and click SCVMM. Save the aws-systems-manager-1.5.zip file to your instance.
2. Open the VMM console.
3. In the navigation pane, click Settings and then click Console Add-Ins.
4. On the ribbon, click Import Console Add-in.
5. On the Select an Add-in page, click Browse and select the aws-systems-manager-1.5.zip file for the add-in that you downloaded.
6. Ignore any warnings that there are assemblies in the add-in that are not signed by a trusted authority. Select Continue installing this add-in anyway and then click Next.
7. On the Summary page, click Finish.
8. When the add-in is imported, the status of the job is Completed. You can close the Jobs window.

Provide Your AWS Credentials

When you use the Systems Manager for the first time, you must provide your AWS credentials. Your access keys identify you to AWS. There are two types of access keys: access key IDs (for example, AKIAIOSFODNN7EXAMPLE) and secret access keys (for example, wJalrXUtntFEMI/K7MDENG/bPxRfYvYqTEgE0=). You should have stored your access keys in a safe place when you received them.

To provide your AWS credentials

1. Open the VMM console.
2. In the navigation pane, click VMs and Services.
3. On the ribbon, click Amazon EC2.
4. On the Credentials tab, specify your AWS credentials, select a default region, and then click Save.
Managing EC2 Instances Using AWS Systems Manager for Microsoft SCVMM

After you log in to the Systems Manager console using your AWS credentials, you can manage your EC2 instances.

Tasks

- Creating an EC2 Instance (p. 983)
- Viewing Your Instances (p. 985)
- Connecting to Your Instance (p. 985)
- Rebooting Your Instance (p. 986)
- Stopping Your Instance (p. 986)
- Starting Your Instance (p. 986)
- Terminating Your Instance (p. 986)

Creating an EC2 Instance

The permissions that you've been granted by your administrator determine whether you can create instances.

Prerequisites

- A virtual private cloud (VPC) with a subnet in the Availability Zone where you'll launch the instance.
  For more information about creating a VPC, see the Amazon VPC Getting Started Guide.

To create an EC2 instance

1. Open SCVMM.
2. On the ribbon, click Create Amazon EC2 Instance.
3. Complete the Create Amazon EC2 Instance dialog box as follows:
   a. Select a region for your instance. By default, we select the region that you configured as your default region.
   b. Select a template (known as an AMI) for your instance. To use an AMI provided by Amazon, select Windows or Linux and then select an AMI from Image. To use an AMI that you created, select My images and then select the AMI from Image.
   c. Select an instance type for the instance. First, select one of the latest instance families from Family, and then select an instance type from Instance type. To include previous generation instance families in the list, select Show previous generations. For more information, see Amazon EC2 Instances and Previous Generation Instances.
   d. Create or select a key pair. To create a key pair, select Create a new key pair from Key pair name and enter a name for the key pair in the highlighted field (for example, my-key-pair).
   e. (Optional) Under Advanced settings, specify a display name for the instance.
   f. (Optional) Under Advanced settings, select a VPC from Network (VPC). Note that this list includes all VPCs for the region, including VPCs created using the Amazon VPC console and the default VPC (if it exists). If you have a default VPC in this region, we select it by default. If the
text is "There is no VPC available for launch or import operations in this region", then you must create a VPC in this region using the Amazon VPC console.

g. (Optional) Under **Advanced settings**, select a subnet from **Subnet**. Note that this list includes all subnets for the selected VPC, including any default subnets. If this list is empty, you must add a subnet to the VPC using the Amazon VPC console, or select a different VPC. Otherwise, we select a subnet for you.

h. (Optional) Under **Advanced settings**, create a security group or select one or more security groups. If you select **Create default security group**, we create a security group that grants RDP and SSH access to everyone, which you can modify using the Amazon EC2 or Amazon VPC console. You can enter a name for this security group in the **Group name** box.

i. (Optional) Under **Advanced settings**, select an IAM role. If this list is empty, you can create a role using the IAM console.

4. Click **Create**. If you are creating a key pair, you are prompted to save the .pem file. Save this file in a secure place; you'll need it to log in to your instance. You'll receive confirmation that the instance has launched. Click **Close**.

After you've created your instance, it appears in the list of instances for the region in which you launched it. Initially, the status of the instance is **pending**. After the status changes to **running**, your instance is ready for use.

You can manage the lifecycle of your instance using Systems Manager, as described on this page. To perform other tasks, such as the following, you must use the AWS Management Console:

- **Attach an Amazon EBS volume to your instance** (p. 803)
- **Associate an Elastic IP address with your instance** (p. 677)
- **Enable termination protection** (p. 390)
Viewing Your Instances

The permissions that your administrator grants you determine whether you can view instances and get detailed information about them.

To view your instances and get detailed information

1. Open the AWS Systems Manager console.
2. From the region list, select a region.
3. From the list of instances, select one or more instances.
4. In the lower pane, click the down arrow next to each instance to view detailed information about the instance.

Connecting to Your Instance

You can log in to an EC2 instance if you have the private key (.pem file) for the key pair that was specified when launching the instance. The tool that you'll use to connect to your instance depends on whether the instance is a Windows instance or a Linux instance.

To connect to a Windows EC2 instance

1. Open AWS Systems Manager.
2. From the list of instances, select the instance, right-click, and then click Retrieve Windows Password.
3. In the Retrieve Default Windows Administrator Password dialog box, click Browse. Select the private key file for the key pair and then click Open.
4. Click Decrypt Password. Save the password or copy it to the clipboard.
5. Select the instance, right-click, and then click Connect via RDP. When prompted for credentials, use the name of the administrator account and the password that you saved in the previous step.
6. Because the certificate is self-signed, you might get a warning that the security certificate is not from a trusted certifying authority. Click Yes to continue.

If the connection fails, see Troubleshooting Windows Instances in the Amazon EC2 User Guide for Windows Instances.

To connect to a Linux EC2 instance

1. Open AWS Systems Manager.
2. From the list of instances, select the instance.
3. In the lower pane, click the down arrow next to the instance ID to view detailed information about the instance.
4. Locate the public DNS name. You’ll need this information to connect to your instance.
5. Connect to the instance using PuTTY. For step-by-step instructions, see Connect to Your Linux Instance from Windows Using PuTTY in the Amazon EC2 User Guide for Linux Instances.

Rebooting Your Instance

The permissions that you've been granted by your administrator determine whether you can reboot instances.

To reboot your instance
1. Open AWS Systems Manager.
2. From the list of instances, select the instance.
3. Right-click the instance, and then click Reset (Reboot).
4. When prompted for confirmation, click Yes.

Stopping Your Instance

The permissions that you've been granted by your administrator determine whether you can stop instances.

To stop your instance
1. Open AWS Systems Manager.
2. From the list of instances, select the instance.
3. Right-click the instance, and then click Shut Down (Stop).
4. When prompted for confirmation, click Yes.

Starting Your Instance

The permissions that you've been granted by your administrator determine whether you can start instances.

To start your instance
1. Open AWS Systems Manager.
2. From the list of instances, select the instance.
3. Right-click the instance, and then click Power On (Start).
4. When prompted for confirmation, click Yes.

If you get a quota error when you try to start an instance, you have reached your concurrent running instance limit. The default limit for your AWS account is 20. If you need additional running instances, complete the form at Request to Increase Amazon EC2 Instance Limit.

Terminating Your Instance

The permissions that you've been granted by your administrator determine whether you can terminate instances.

To terminate your instance
1. Open AWS Systems Manager.
Importing Your VM

2. From the list of instances, select the instance.
3. Right-click the instance, and then click **Delete (Terminate)**.
4. When prompted for confirmation, click **Yes**.

Importing Your Virtual Machine Using AWS Systems Manager for Microsoft SCVMM

You can launch an EC2 instance from a virtual machine that you import from SCVMM to Amazon EC2.

**Important**
You can't import Linux virtual machines from SCVMM to Amazon EC2.

**Contents**
- Prerequisites (p. 987)
- Importing Your Virtual Machine (p. 987)
- Checking the Import Task Status (p. 988)
- Backing Up Your Imported Instance (p. 989)

**Prerequisites**

- Ensure that your VM is ready. For more information, see **Prepare Your VM** in the **VM Import/Export User Guide**.
- In AWS Systems Manager, click **Configuration**, select the **VM Import** tab, and review the following settings:
  - **S3 bucket prefix**: We create a bucket for disk images to be uploaded before they are imported. The name of the bucket starts with the prefix listed here and includes the region (for example, us-east-2). To delete the disk images after they are imported, select **Clean up S3 bucket after import**.
  - **VM image export path**: A location for the disk images exported from the VM. To delete the disk images after they are imported, select **Clean up export path after import**.
  - **Alternate Hyper-V PowerShell module path**: The location of the Hyper-V PowerShell module, if it's not installed in the standard location. For more information, see **Installing the Hyper-V Management Tools** in the Microsoft TechNet Library.

**Importing Your Virtual Machine**

The permissions that you've been granted by your administrator determine whether you can import HyperV Windows virtual machines from SCVMM to AWS.

**To import your virtual machine**

1. Open SCVMM.
2. On the ribbon, click **VMs**. Select your virtual machine from the list.
3. On the ribbon, click **Import VM to Amazon EC2**.
4. Complete the **Import Virtual Machine** dialog box as follows:
   - Select a region for the instance. By default, we select the region that you configured as your default region.
   - Select an instance type for the instance. First, select one of the latest instance families from **Family**, and then select an instance type from **Instance type**. To include previous generation
instance families in the list, select **Show previous generations**. For more information, see *Amazon EC2 Instances* and *Previous Generation Instances*.

- **c. Select a VPC from Network (VPC).** Note that this list includes all VPCs for the region, including VPCs created using the Amazon VPC console and the default VPC (if it exists). If you have a default VPC in this region, we select it by default. If the text is "There is no VPC available for launch or import operations in this region", then you must create a VPC in this region using the Amazon VPC console.

- **d. Select a subnet from Subnet.** Note that this list includes all subnets for the selected VPC, including any default subnets. If this list is empty, you must add a subnet to the VPC using the Amazon VPC console, or select a different VPC. Otherwise, we select a subnet for you.

5. Click **Import**. If you haven't specified the required information in the **VM Import** tab, you'll receive an error asking you to provide the required information. Otherwise, you'll receive confirmation that the import task has started. Click **Close**.

### Checking the Import Task Status

The import task can take several hours to complete. To view the current status, open AWS System Manager and click **Notifications**.

You'll receive the following notifications as the import task progresses:

- **Import VM: Created Import VM Task**
- **Import VM: Export VM Disk Image Done**
- **Import VM: Upload to S3**
- **Import VM: Image Conversion Starting**
- **Import VM: Image Conversion Done**
- **Import VM: Import Complete**

**Note** that you'll receive the **Import VM: Upload to S3**, **Import VM: Image Conversion Starting**, and **Import VM: Image Conversion Done** notifications for each disk image converted.
If the import task fails, you'll receive the notification Import VM: Import Failed. For more information about troubleshooting issues with import tasks, see Errors Importing a VM (p. 990).

**Backing Up Your Imported Instance**

After the import operation completes, the instance runs until it is terminated. If your instance is terminated, you can't connect to or recover the instance. To ensure that you can start a new instance with the same software as an imported instance if needed, create an Amazon Machine Image (AMI) from the imported instance. For more information, see Creating a Custom Windows AMI (p. 64).

**Troubleshooting AWS Systems Manager for Microsoft SCVMM**

The following are common errors and troubleshooting steps.

**Contents**

- Error: Add-in cannot be installed (p. 989)
- Installation Errors (p. 989)
- Checking the Log File (p. 990)
- Errors Importing a VM (p. 990)
- Uninstalling the Add-In (p. 990)

**Error: Add-in cannot be installed**

If you receive the following error, try installing KB2918659 on the computer running the VMM console. For more information, see Description of System Center 2012 SP1 Update Rollup 5. Note that you don't need to install all the updates listed in this article to address this issue, just KB2918659.

Add-in cannot be installed
The assembly "Amazon.Scvmm.Addin" referenced to by add-in component "AWS Systems Manager for Microsoft SCVMM" could not be found in the add-in package. This could be due to the following reasons:
1. The assembly was not included with the add-in package.
2. The AssemblyName attribute for the add-in does not match the name of the add-in assembly.
3. The assembly file is corrupt and cannot be loaded.

**Installation Errors**

If you receive one of the following errors during installation, it is likely due to an issue with SCVMM:

Could not update managed code add-in pipeline due to the following error:
Access to the path 'C:\Program Files\Microsoft System Center 2012\Virtual Machine Manager \Bin\AddInPipeline\PipelineSegments.store' is denied.

Could not update managed code add-in pipeline due to the following error:
The required folder 'C:\Program Files\Microsoft System Center 2012\Virtual Machine Manager \Bin\AddInPipeline\HostSideAdapters' does not exist.
Add-in cannot be installed
The assembly "Microsoft.SystemCenter.VirtualMachineManager.UIAddIns.dll" referenced by the
add-in assembly "Amazon.Scvmm.AddIn" could not be found in the add-in package. Make sure
that this assembly was included with the add-in package.

Try one of the following steps to work around this issue:

- Grant authenticated users permission to read and execute the C:\Program Files\Microsoft
  System Center 2012\Virtual Machine Manager\Bin\AddInPipeline folder. In Windows
  Explorer, right-click the folder, select Properties, and then select the Security tab.
- Close the SCVMM console and start it one time as an administrator. From the Start menu, locate
  SCVMM, right-click, and then select Run as administrator.

Checking the Log File

If you have a problem using the add-in, check the generated log file, %APPDATA%\Amazon\SCVMM
\ec2addin.log, for useful information.

Errors Importing a VM

The log file, %APPDATA%\Amazon\SCVMM\ec2addin.log, contains detailed information about the
status of an import task. The following are common errors that you might see in the log file when you
import your VM from SCVMM to Amazon EC2.

**Error: Unable to extract Hyper-V VirtualMachine object**

Solution: Configure the path to the Hyper-V PowerShell module.

**Error: You do not have permission to perform the operation**

This error usually occurs when Hyper-V can't save the VM image into the configured path. To resolve this
issue, do the following.

1. Create a directory on the Hyper-V server. For example: C:vmimages.
2. Share the directory you just created in Hyper-V. Any user running SCVMM should be given access to
   the directory.
3. In the plugin, set the export path to \hyperv\vmimages.
4. Perform the export.

The image will be exported to a local directory on the Hyper-V server. The SCVMM plugin will pull it from
Hyper-V, and upload into Amazon S3.

Uninstalling the Add-In

If you need to uninstall the add-in, use the following procedure.

**To uninstall the add-in**

1. Open the VMM console.
2. Select the Settings workspace, and then click Console Add-Ins.
3. Select AWS Systems Manager for Microsoft SCVMM.
4. On the ribbon, click Remove.
5. When prompted for confirmation, click Yes.
If you reinstall the add-in after uninstalling it and receive the following error, delete the path as suggested by the error message.

Error (27301)
There was an error while installing the add-in. Please ensure that the following path does not exist and then try the installation again.

C:\Program Files\Microsoft System Center 2012\Virtual Machine Manager\Bin\AddInPipeline\AddIns\EC2WINDOWS...
AWS Management Pack for Microsoft System Center

Amazon Web Services (AWS) offers a complete set of infrastructure and application services for running almost anything in the cloud—from enterprise applications and big data projects to social games and mobile apps. The AWS Management Pack for Microsoft System Center provides availability and performance monitoring capabilities for your applications running in AWS.

The AWS Management Pack allows Microsoft System Center Operations Manager to access your AWS resources (such as instances and volumes), so that it can collect performance data and monitor your AWS resources. The AWS Management Pack is an extension to System Center Operations Manager. There are two versions of the AWS Management Pack: one for System Center 2012 — Operations Manager and another for System Center Operations Manager 2007 R2.

The AWS Management Pack uses Amazon CloudWatch metrics and alarms to monitor your AWS resources. Amazon CloudWatch metrics appear in Microsoft System Center as performance counters and Amazon CloudWatch alarms appear as alerts.

You can monitor the following resources:

- EC2 instances
- EBS volumes
- ELB load balancers
- Amazon EC2 Auto Scaling groups and Availability Zones
- Elastic Beanstalk applications
- CloudFormation stacks
- CloudWatch Alarms
- CloudWatch Custom Metrics

Contents

- Overview of AWS Management Pack for System Center 2012 (p. 992)
- Overview of AWS Management Pack for System Center 2007 R2 (p. 994)
- Downloading the AWS Management Pack (p. 995)
- Deploying the AWS Management Pack (p. 996)
- Using the AWS Management Pack (p. 1005)
- Upgrading the AWS Management Pack (p. 1025)
- Uninstalling the AWS Management Pack (p. 1026)
- Troubleshooting the AWS Management Pack (p. 1027)

Overview of AWS Management Pack for System Center 2012

The AWS Management Pack for System Center 2012 — Operations Manager uses a resource pool that contains one or more management servers to discover and monitor your AWS resources. You can add management servers to the pool as you increase the number of AWS resources that you use.
The following diagram shows the main components of AWS Management Pack.

<table>
<thead>
<tr>
<th>Item</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Operations Manager infrastructure</td>
<td>One or more management servers and their dependencies, such as Microsoft SQL Server and a Microsoft Active Directory domain. These servers can either be deployed on-premises or in the AWS cloud; both scenarios are supported.</td>
</tr>
<tr>
<td>2</td>
<td>Resource pool</td>
<td>One or more management servers used for communicating with AWS using the AWS SDK for .NET. These servers must have Internet connectivity.</td>
</tr>
<tr>
<td>3</td>
<td>AWS credentials</td>
<td>An access key ID and a secret access key used by the management servers to make AWS API calls. You must specify these credentials while you configure the AWS Management Pack. We recommend that you create an IAM user with read-only privileges and use those credentials. For more information about creating an IAM user, see Adding a New User to Your AWS Account in the IAM User Guide.</td>
</tr>
<tr>
<td>4</td>
<td>EC2 instances</td>
<td>Virtual computers running in the AWS cloud. Some instances might have the Operations Manager Agent installed, others might not. When you install Operations Manager Agent you can see the operating system and application health apart from the instance health.</td>
</tr>
</tbody>
</table>
Overview of AWS Management Pack for System Center 2007 R2

The AWS Management Pack for System Center Operations Manager 2007 R2 uses a designated computer that connects to your System Center environment and has Internet access, called a **watcher node**, to call AWS APIs to remotely discover and collect information about your AWS resources.

The following diagram shows the main components of AWS Management Pack.

<table>
<thead>
<tr>
<th>Item</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Operations Manager infrastructure</td>
<td>One or more management servers and their dependencies, such as Microsoft SQL Server and a Microsoft Active Directory domain. These servers can either be deployed on-premises or in the AWS cloud; both scenarios are supported.</td>
</tr>
<tr>
<td>2</td>
<td>Watcher node</td>
<td>A designated agent-managed computer used for communicating with AWS using the AWS SDK for .NET. It can either be deployed on-premises or in the AWS cloud, but it must be an agent-managed computer, and it must have Internet connectivity. You can use exactly one watcher node to monitor an AWS account. However, one watcher node can monitor multiple AWS accounts. For more information about setting up a watcher node, see Deploying Windows Agents in the Microsoft System Center documentation.</td>
</tr>
<tr>
<td>3</td>
<td>AWS credentials</td>
<td>An access key ID and a secret access key used by the watcher node to make AWS API calls. You must specify these credentials while you configure the AWS Management Pack. We recommend that you create an IAM user with read-only access.</td>
</tr>
</tbody>
</table>
### Downloading the AWS Management Pack

To get started, download the AWS Management Pack. The AWS Management Pack is free. You might incur charges for Amazon CloudWatch, depending on how you configure monitoring or how many AWS resources you monitor.

#### System Center 2012

Before you download the AWS Management Pack, ensure that your systems meet the following system requirements and prerequisites.

**System Requirements**

- System Center Operations Manager 2012 R2 or System Center Operations Manager 2012 SP1
- Cumulative Update 1 or later. You must deploy the update to the management servers monitoring AWS resources, as well as agents running the watcher nodes and agents to be monitored by the AWS Management Pack. We recommend that you deploy the latest available Operations Manager updates on all computers monitoring AWS resources.
- Microsoft.Unix.Library MP version 7.3.2026.0 or later

**Prerequisites**

- Your data center must have at least one management server configured with Internet connectivity. The management servers must have the Microsoft .NET Framework version 4.5 or later and PowerShell 2.0 or later installed.
- The action account for the management server must have local administrator privileges on the management server.

**To download the AWS Management Pack**

1. On the [AWS Add-Ins for Microsoft System Center](https://aws.amazon.com) website, click **SCOM 2012**.
2. Save `AWS-SCOM-MP-2.5.zip` to your computer and unzip it.

Continue with [Deploying the AWS Management Pack](p. 996).

#### System Center 2007 R2

Before you download the AWS Management Pack, ensure that your systems meet the following system requirements and prerequisites.

<table>
<thead>
<tr>
<th>Item</th>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>privileges and use those credentials. For more information about creating an IAM user, see <a href="https://docs.aws.amazon.com/IAM/latest/UserGuide/">Adding a New User to Your AWS Account</a> in the <em>IAM User Guide</em>.</td>
</tr>
<tr>
<td>4</td>
<td>EC2 instances</td>
<td>Virtual computers running in the AWS cloud. Some instances might have the Operations Manager Agent installed, others might not. When you install the Operations Manager Agent you can see the operating system and application health apart from the instance health.</td>
</tr>
</tbody>
</table>
System Requirements

- System Center Operations Manager 2007 R2
- Microsoft.Unix.Library MP version 6.1.7000.256 or later

Prerequisites

- Your data center must have an agent-managed computer with Internet connectivity that you designate as the watcher node. The watcher node must have the following Agent Proxy option enabled: **Allow this agent to act as a proxy and discover managed objects on other computers.** The watcher node must have the Microsoft .NET Framework version 3.5.1 or later and PowerShell 2.0 or later installed.
- The action account for the watcher node must have local administrator privileges on the watcher node.
- You must ensure that your watcher node has the agent installed, has Internet access, and can communicate with the management servers in your data center. For more information, see Deploying Windows Agents in the Microsoft System Center documentation.

To download the AWS Management Pack

1. On the AWS Add-Ins for Microsoft System Center website, click **SCOM 2007**.
2. Save **AWS-MP-Setup-2.5.msi** to your computer.

Continue with Deploying the AWS Management Pack (p. 996).

Deploying the AWS Management Pack

Before you can deploy the AWS Management Pack, you must download it. For more information, see Downloading the AWS Management Pack (p. 995).

Tasks

- Step 1: Installing the AWS Management Pack (p. 996)
- Step 2: Configuring the Watcher Node (p. 998)
- Step 3: Create an AWS Run As Account (p. 998)
- Step 4: Run the Add Monitoring Wizard (p. 1001)
- Step 5: Configure Ports and Endpoints (p. 1005)

Step 1: Installing the AWS Management Pack

After you download the AWS Management Pack, you must configure it to monitor one or more AWS accounts.

System Center 2012

To install the AWS Management Pack

1. In the Operations console, on the Go menu, click **Administration**, and then click **Management Packs**.
2. In the Actions pane, click **Import Management Packs**.
3. On the **Select Management Packs** page, click **Add**, and then click **Add from disk**.

4. In the **Select Management Packs to import** dialog box, select the Amazon.AmazonWebServices.mpb file from the location where you downloaded it, and then click **Open**.

5. On the **Select Management Packs** page, under **Import list**, select the **Amazon Web Services** management pack, and then click **Install**.

   **Note**
   System Center Operations Manager doesn’t import any management packs in the **Import** list that display an **Error** icon.

6. The **Import Management Packs** page shows the progress for the import process. If a problem occurs, select the management pack in the list to view the status details. Click **Close**.

### System Center 2007 R2

#### To install the AWS Management Pack

The management pack is distributed as a Microsoft System Installer file, AWS-MP-Setup.msi. It contains the required DLLs for the watcher node, root management server, and Operations console, as well as the Amazon.AmazonWebServices.mp file.

1. Run AWS-MP-Setup.msi.

   **Note**
   If your root management server, Operations console, and watcher node are on different computers, you must run the installer on each computer.

2. On the **Welcome to the Amazon Web Services Management Pack Setup Wizard** screen, click **Next**.

3. On the **End-User License Agreement** screen, read the license agreement, and, if you accept the terms, select the **I accept the terms in the License Agreement** check box, and then click **Next**.

4. On the **Custom Setup** screen, select the features you want to install, and then click **Next**.

   **Operations Console**
   Installs Amazon.AmazonWebServices.UI.Pages.dll and registers it in the Global Assembly Cache (GAC), and then installs Amazon.AmazonWebServices.mp.

   **Root Management Server**
   Installs Amazon.AmazonWebServices.Modules.dll, Amazon.AmazonWebServices.SCOM.SDK.dll and the AWS SDK for .NET (AWSSDK.dll), and then registers them in the GAC.

   **AWS Watcher Node**
   Installs Amazon.AmazonWebServices.Modules.dll and Amazon.AmazonWebServices.SCOM.SDK.dll, and then installs the AWS SDK for .NET (AWSSDK.dll) and registers it in the GAC.

5. On the **Ready to install Amazon Web Services Management Pack** screen, click **Install**.

6. On the **Completed the Amazon Web Services Management Pack Setup Wizard** screen, click **Finish**.

   **Note**
   The required DLLs are copied and registered in the GAC, and the management pack file (*.mp) is copied to the Program Files (x86)/Amazon Web Services Management Pack folder on the computer running the Operations console. Next, you must import the management pack into System Center.

7. In the Operations console, on the **Go** menu, click **Administration**, and then click **Management Packs**.
8. In the Actions pane, click Import Management Packs.
9. On the Select Management Packs page, click Add, and then click Add from disk.
10. In the Select Management Packs to import dialog box, change the directory to C:\Program Files (x86)\Amazon Web Services Management Pack, select the Amazon.AmazonWebServices.mp file, and then click Open.
11. On the Select Management Packs page, under Import list, select the Amazon Web Services management pack, and then click Install.

   Note
   System Center Operations Manager doesn't import any management packs in the Import list that display an Error icon.

12. The Import Management Packs page shows the progress for the import process. If a problem occurs, select the management pack in the list to view the status details. Click Close.

Step 2: Configuring the Watcher Node

On System Center Operations Manager 2007 R2, the watcher node runs discoveries that go beyond the watcher node computer, so you must enable the proxy agent option on the watcher node. The proxy agent allows those discoveries to access the objects on other computers.

   Note
   If your system is configured with a large number of resources, we recommend that you configure one management server as a Watcher Node. Having a separate Watcher Node management server can improve performance.

If you’re using System Center 2012 — Operations Manager, you can skip this step.

To enable the proxy agent on System Center Operations Manager 2007 R2

1. In the Operations console, on the Go menu, click Administration.
2. In the Administration workspace, under Device Management, click Agent Managed.
3. In the Agent Managed list, right-click the watcher node, and then click Properties.
4. In the Agent Properties dialog box, click the Security tab, select Allow this agent to act as proxy and discover managed objects on other computers, and then click OK.

Step 3: Create an AWS Run As Account

You must set up credentials that grant AWS Management Pack access to your AWS resources.

To create an AWS Run As account

1. We recommend that you create an IAM user with the minimum access rights required (for example, the ReadOnlyAccess AWS managed policy works in most cases). You’ll need the access keys (access key ID and secret access key) for this user to complete this procedure. For more information, see Administering Access Keys for IAM Users in the IAM User Guide.
2. In the Operations console, on the Go menu, click Administration.
3. In the Administration workspace, expand the Run As Configuration node, and then select Accounts.
4. Right-click the Accounts pane, and then click Create Run As Account.
5. In the Create Run As Account Wizard, on the General Properties page, in the Run As account type list, select Basic Authentication.
6. Enter a display name (for example, "My IAM Account") and a description, and then click Next.
7. On the **Credentials** page, enter the access key ID in the **Account name** box and the secret access key in the **Password** box, and then click **Next**.

8. On the **Distribution Security** page, select *More secure - I want to manually select the computers to which the credentials will be distributed*, and then click **Create**.
9. Click **Close**.

10. In the list of accounts, select the account that you just created.

11. In the **Actions** pane, click **Properties**.

12. In the **Properties** dialog box, verify that the **More Secure** option is selected and that all management servers to be used to monitor your AWS resources are listed.
Step 4: Run the Add Monitoring Wizard

You can configure the AWS Management Pack to monitor a particular AWS account by using the Add Monitoring Wizard, which is available in the Authoring workspace of the Operations console. This wizard creates a management pack that contains the settings for the AWS account to monitor. You must run this wizard to monitor each AWS account. For example, if you want to monitor two AWS accounts, you must run the wizard twice.

System Center 2012

To run the Add Monitoring Wizard on System Center 2012 — Operations Manager

1. In the Operations console, on the Go menu, click Authoring.
2. In the Authoring workspace, expand the Management Pack Templates node, right-click Amazon Web Services, and then click Add Monitoring Wizard.
3. In the Add Monitoring Wizard, in the Select the monitoring type list, select Amazon Web Services, and then click Next.
4. On the General Properties page, in the Name box, enter a name (for example, “My AWS Resources”). In the Description box, enter a description.
5. In the Select destination management pack list, select an existing management pack (or click New to create one) where you want to save the settings. Click Next.

By default, when you create a management pack object, disable a rule or monitor, or create an override, Operations Manager saves the setting to the default management pack. As a best practice, you should create a separate management pack for each sealed management pack that you want to customize, instead of saving your customized settings to the default management pack.

6. The AWS Management Pack automatically creates a resource pool and adds the management servers to it. To control server membership, make the following changes:
a. Click **Administration** on the **Go** menu.
b. Click the **Resource Pools** node.

c. Right-click the **AWS Resource Pool** in the **Resource Pools** pane and select **Manual Membership**.

d. Right-click the **AWS Resource Pool** in the **Resource Pools** pane and select **Properties**.
e. On the **Pool Membership** page, remove the management servers that should not monitor AWS resources.

7. After the AWS Management Pack is configured, it shows up as a sub-folder of the **Amazon Web Services** folder in the **Monitoring** workspace of the Operations console.
System Center 2007 R2

To run the Add Monitoring Wizard on System Center Operations Manager 2007

1. In the Operations console, on the Go menu, click Authoring.

2. In the Authoring workspace, expand the Management Pack Templates node, right-click Amazon Web Services, and then click Add Monitoring Wizard.

3. In the Add Monitoring Wizard, in the Select the monitoring type list, select Amazon Web Services, and then click Next.

4. On the General Properties page, in the Name box, enter a name (for example, “My AWS Resources”). In the Description box, enter a description.

5. In the Select destination management pack drop-down list, select an existing management pack (or click New to create a new one) where you want to save the settings. Click Next.
By default, when you create a management pack object, disable a rule or monitor, or create an override, Operations Manager saves the setting to the default management pack. As a best practice, you should create a separate management pack for each sealed management pack that you want to customize, instead of saving your customized settings to the default management pack.

6. On the **Watcher Node Configuration** page, in the **Watcher Node** list, select an agent-managed computer to act as the watcher node.

7. In the **Select AWS Run As account** drop-down list, select the Run As account that you created earlier, and then click **Create**.

8. After the AWS Management Pack is configured, it first discovers the watcher node. To verify that the watcher node was discovered successfully, navigate to the **Monitoring** workspace in the Operations console. You should see a new **Amazon Web Services** folder and an **Amazon Watcher Nodes** subfolder under it. This subfolder displays the watcher nodes. The AWS Management Pack automatically checks and monitors the watcher node connectivity to AWS. When the watcher node is discovered, it shows up in this list. When the watcher node is ready, its state changes to **Healthy**.

**Note**

To establish connectivity with AWS, the AWS Management Pack requires that you deploy the AWS SDK for .NET, modules, and scripts to the watcher node. This can take about ten minutes. If the watcher node doesn't appear, or if you see the state as **Not Monitored**, verify your Internet connectivity and IAM permissions. For more information, see **Troubleshooting the AWS Management Pack** (p. 1027).
After the watcher node is discovered, dependent discoveries are triggered, and the AWS resources are added to the Monitoring workspace of the Operations console.

The discovery of AWS resources should finish within twenty minutes. This process can take more time, based on your Operations Manager environment, your AWS environment, the load on the management server, and the load on the watcher node. For more information, see Troubleshooting the AWS Management Pack (p. 1027).

Step 5: Configure Ports and Endpoints

The AWS Management Pack for Microsoft System Center must be able to communicate with AWS services to monitor the performance of those services and provide alerts in System Center. For monitoring to succeed, you must configure the firewall on the Management Pack servers to allow outbound HTTP calls on ports 80 and 443 to the AWS endpoints for the following services.

This enables monitoring for the following AWS services:

- Amazon Elastic Compute Cloud (EC2)
- Elastic Load Balancing
- Amazon EC2 Auto Scaling
- AWS Elastic Beanstalk
- Amazon CloudWatch
- AWS CloudFormation

The AWS Management Pack uses the public APIs in the AWS SDK for .NET to retrieve information from these services over ports 80 and 443. Log on to each server and enable outbound firewall rules for ports 80 and 443.

If your firewall application supports more detailed settings you can configure specific endpoints for each service. An endpoint is a URL that is the entry point for a web service. For example, ec2.us-west-2.amazonaws.com is an entry point for the Amazon EC2 service. To configure endpoints on your firewall, locate the specific endpoint URLs for the AWS services you are running and specify those endpoints in your firewall application.

Using the AWS Management Pack

You can use the AWS Management Pack to monitor the health of your AWS resources.
Views

The AWS Management Pack provides the following views, which are displayed in the Monitoring workspace of the Operations console.

Views

- EC2 Instances (p. 1006)
- Amazon EBS Volumes (p. 1008)
- Elastic Load Balancers (p. 1010)
- AWS Elastic Beanstalk Applications (p. 1012)
- AWS CloudFormation Stacks (p. 1014)
- Amazon Performance Views (p. 1016)
- Amazon CloudWatch Metric Alarms (p. 1017)
- AWS Alerts (p. 1018)
- Watcher Nodes (System Center Operations Manager 2007 R2) (p. 1019)

EC2 Instances

View the health state of the EC2 instances for a particular AWS account, from all Availability Zones and regions. The view also includes EC2 instances running in a virtual private cloud (VPC). The AWS Management Pack retrieves tags, so you can search and filter the list using those tags.
When you select an EC2 instance, you can perform instance health tasks:

- **Open Amazon Console**: Launches the AWS Management Console in a web browser.
- **Open RDP to Amazon EC2 Instance**: Opens an RDP connection to the selected Windows instance.
- **Reboot Amazon EC2 Instance**: Reboots the selected EC2 instance.
- **Start Amazon EC2 Instance**: Starts the selected EC2 instance.
- **Stop Amazon EC2 Instance**: Stops the selected EC2 instance.

**EC2 Instances Diagram View**

Shows the relationship of an instance with other components.
Amazon EBS Volumes

Shows the health state of all the Amazon EBS volumes for a particular AWS account from all Availability Zones and regions.
Amazon EBS Volumes Diagram View

Shows an Amazon EBS volume and any associated alarms. The following illustration shows an example:
Elastic Load Balancers

Shows the health state of all the load balancers for a particular AWS account from all regions.
Elastic Load Balancing Diagram View

Shows the Elastic Load Balancing relationship with other components. The following illustration shows an example:
AWS Elastic Beanstalk Applications

Shows the state of all discovered AWS Elastic Beanstalk applications.
AWS Elastic Beanstalk Applications Diagram View

Shows the AWS Elastic Beanstalk application, application environment, application configuration, and application resources objects.
AWS CloudFormation Stacks

Shows the health state of all the AWS CloudFormation stacks for a particular AWS account from all regions.
AWS CloudFormation Stacks Diagram View

Shows the AWS CloudFormation stack relationship with other components. An AWS CloudFormation stack might contain Amazon EC2 or Elastic Load Balancing resources. The following illustration shows an example:
Amazon Performance Views

Shows the Amazon CloudWatch metrics for Amazon EC2, Amazon EBS, and Elastic Load Balancing, custom metrics, and metrics created from CloudWatch alarms. In addition, there are separate performance views for each resource. The Other Metrics performance view contains custom metrics, and metrics created from CloudWatch alarms. For more information about these metrics, see the CloudWatch Metrics, Namespaces, and Dimensions Reference in the Amazon CloudWatch Developer Guide. The following illustration shows an example.
Amazon CloudWatch Metric Alarms

Shows Amazon CloudWatch alarms related to the discovered AWS resources.
AWS Alerts

Shows the alerts that the AWS management pack produces when the health of an object is in a critical state.
Watcher Nodes (System Center Operations Manager 2007 R2)

View the health state of the watcher nodes across all of the AWS accounts that are being monitored. A Healthy state means that the watcher node is configured correctly and can communicate with AWS.
Discoveries

Discoveries are the AWS resources that are monitored by the AWS Management Pack. The AWS Management Pack discovers the following objects:

- Amazon EC2 instances
- EBS volumes
- ELB load balancers
- AWS CloudFormation stacks
- Amazon CloudWatch alarms
- AWS Elastic Beanstalk applications
- Amazon EC2 Auto Scaling groups and Availability Zones

Amazon CloudWatch metrics are generated for the following resources:

- Amazon EC2 instance
- EBS volume
- Elastic Load Balancing
- Custom Amazon CloudWatch metrics
- Metrics from existing Amazon CloudWatch alarms

For Amazon CloudWatch metrics discovery, the following guidelines apply:

- AWS CloudFormation stacks do not have any default Amazon CloudWatch metrics.
- Stopped Amazon EC2 instances or unused Amazon EBS volumes do not generate data for their default Amazon CloudWatch metrics.
- After starting an Amazon EC2 instance, it can take up to 30 minutes for the Amazon CloudWatch metrics to appear in Operations Manager.
• Amazon CloudWatch retains the monitoring data for two weeks, even if your AWS resources have been terminated. This data appears in Operations Manager.

• An existing Amazon CloudWatch alarm for a resource that is not supported will create a metric and be associated with the Amazon CloudWatch alarm. These metric can be viewed in the Other Metrics performance view.

The AWS Management Pack also discovers the following relationships:

• AWS CloudFormation stack and its Elastic Load Balancing or Amazon EC2 resources
• Elastic Load Balancing load balancer and its EC2 instances
• Amazon EC2 instance and its EBS volumes
• Amazon EC2 instance and its operating system
• AWS Elastic Beanstalk application and its environment, configuration, and resources

The AWS Management Pack automatically discovers the relationship between an EC2 instance and the operating system running on it. To discover this relationship, the Operations Manager Agent must be installed and configured on the instance and the corresponding operating system management pack must be imported in Operations Manager.

Discoveries run on the management servers in the resource pool (System Center 2012) or the watcher node (System Center 2007 R2).

<table>
<thead>
<tr>
<th>Discovery</th>
<th>Interval (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Resources Discovery (SCOM 2012)</td>
<td>14400</td>
</tr>
<tr>
<td>Discovers EC2 instances, Amazon EBS volumes, load balancers, and CloudFront stacks.</td>
<td></td>
</tr>
<tr>
<td>AWS Elastic Beanstalk Discovery</td>
<td>14400</td>
</tr>
<tr>
<td>Discovers AWS Elastic Beanstalk and its relationship with environment, resources, and configuration.</td>
<td></td>
</tr>
<tr>
<td>CloudWatch Alarms Discovery</td>
<td>900</td>
</tr>
<tr>
<td>Discovers alarms generated using CloudWatch metrics.</td>
<td></td>
</tr>
<tr>
<td>Custom CloudWatch Metric Discovery</td>
<td>14400</td>
</tr>
<tr>
<td>Discovers custom CloudWatch metrics.</td>
<td></td>
</tr>
<tr>
<td>Watcher Node Discovery (SCOM 2007 R2)</td>
<td>14400</td>
</tr>
<tr>
<td>Targets the root management server and creates the watcher node objects.</td>
<td></td>
</tr>
</tbody>
</table>

**Monitors**

Monitors are used to measure the health of your AWS resources. Monitors run on the management servers in the resource pool (System Center 2012) or the watcher node (System Center 2007 R2).
Rules

Rules create alerts (based on Amazon CloudWatch metrics) and collect data for analysis and reporting.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Interval (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS Resource Discovery Rule (SCOM 2007 R2)</td>
<td>14400</td>
</tr>
<tr>
<td>Targets the watcher node and uses the AWS API to discover objects for the following AWS resources: EC2 instances, EBS volumes, load balancers, and AWS CloudFormation stacks. (CloudWatch metrics or alarms are not discovered). After discovery is complete, view the objects in the Not Monitored state.</td>
<td></td>
</tr>
<tr>
<td>Amazon Elastic Block Store Volume Performance Metrics Data Collection Rule</td>
<td>900</td>
</tr>
<tr>
<td>Amazon EC2 Instance Performance Metrics Data Collection Rule</td>
<td>900</td>
</tr>
<tr>
<td>Elastic Load Balancing Balancing Performance Metrics Data Collection Rule</td>
<td>900</td>
</tr>
<tr>
<td>Custom CloudWatch Metric Data Collection Rule</td>
<td>900</td>
</tr>
</tbody>
</table>

Events

Events report on activities that involve the monitored resources. Events are written to the Operations Manager event log.

<table>
<thead>
<tr>
<th>Event ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4101</td>
<td>Amazon EC2 Instance Discovery (General Discovery) finished</td>
</tr>
<tr>
<td>4102</td>
<td>Elastic Load Balancing Metrics Discovery,</td>
</tr>
<tr>
<td></td>
<td>Amazon EBS Volume Metrics Discovery,</td>
</tr>
<tr>
<td></td>
<td>Amazon EC2 Instance Metrics Discovery finished</td>
</tr>
<tr>
<td>4103</td>
<td>Amazon CloudWatch Metric Alarms Discovery finished</td>
</tr>
</tbody>
</table>
Health Model

The following illustration shows the health model defined by the AWS Management Pack.

The health state for a CloudWatch alarm is rolled up to its corresponding CloudWatch metric. The health state for a CloudWatch metric for Amazon EC2 is rolled up to the EC2 instance. Similarly, the health state for the CloudWatch metrics for Amazon EBS is rolled up to the Amazon EBS volume. The health states for the Amazon EBS volumes used by an EC2 instance are rolled up to the EC2 instance.

When the relationship between an EC2 instance and its operating system has been discovered, the operating system health state is rolled up to the EC2 instance.
The health state of an AWS CloudFormation stack depends on the status of the AWS CloudFormation stack itself and the health states of its resources, namely the load balancers and EC2 instances.

The following table illustrates how the status of the AWS CloudFormation stack corresponds to its health state.

<table>
<thead>
<tr>
<th>Health State</th>
<th>AWS CloudFormation Stack Status</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error</td>
<td>CREATE_FAILED</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DELETE_IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DELETE_FAILED</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UPDATE_ROLLBACK_FAILED</td>
<td>Most likely usable</td>
</tr>
<tr>
<td>Warning</td>
<td>UPDATE_ROLLBACK_IN_PROGRESS</td>
<td>Recovering after some problem</td>
</tr>
<tr>
<td></td>
<td>UPDATE_ROLLBACK_COMPLETE_CLEANUP_IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UPDATE_ROLLBACK_COMPLETE</td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>CREATE_COMPLETE</td>
<td>Usable</td>
</tr>
<tr>
<td></td>
<td>UPDATE_IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UPDATE_COMPLETE_CLEANUP_IN_PROGRESS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UPDATE_COMPLETE</td>
<td></td>
</tr>
</tbody>
</table>

The full health model for an AWS CloudFormation stack is as follows:

**Customizing the AWS Management Pack**

To change the frequency of discoveries, rules, and monitors, you can override the interval time (in seconds).
To change frequency

1. In the Operations Manager toolbar, click Go, and then click Authoring.
2. In the Authoring pane, expand Management Pack Objects and then click the object to change (for example, Object Discoveries, Rules, or Monitors).
3. In the toolbar, click Scope.
4. In the Scope Management Pack Objects dialog box, click View all targets.
5. To limit the scope to Amazon objects, type Amazon in the Look for field.
6. Select the object want to configure and click OK.
7. In the Operations Manager center pane, right-click the object to configure, click Overrides, and then click the type of override you want to configure.
8. Use the Override Properties dialog box to configure different values and settings for objects.

Tip
To disable a discovery, rule, or monitoring object right-click the object to disable in the Operations Manager center pane, click Overrides, and then click Disable the Rule. You might disable rules if, for example, you do not run AWS Elastic Beanstalk applications or use custom Amazon CloudWatch metrics.

For information about creating overrides, see Tuning Monitoring by Using Targeting and Overrides on the Microsoft TechNet website.

For information about creating custom rules and monitors, see Authoring for System Center 2012 - Operations Manager or System Center Operations Manager 2007 R2 Management Pack Authoring Guide on the Microsoft TechNet website.

Upgrading the AWS Management Pack

The procedure that you’ll use to update AWS Management Pack depends on the version of System Center.

System Center 2012

To upgrade the AWS Management Pack

2. In the Operations console, on the Go menu, click Administration, and then click Management Packs.
3. In the Tasks pane, click Import Management Packs.
4. On the Select Management Packs page, click Add, and then click Add from disk.
5. In the Select Management Packs to import dialog box, select the Amazon.AmazonWebServices.mpb file from the location where you downloaded it, and then click Open.
6. On the Select Management Packs page, under Import list, select the Amazon Web Services management pack, and then click Install.

If the Install button is disabled, upgrading to the current version is not supported and you must uninstall the AWS Management Pack before you can install the current version. For more information, see Uninstalling the AWS Management Pack (p. 1026).
System Center 2007 R2

To upgrade the AWS Management Pack

1. On the Management Server, go to the AWS Add-Ins for Microsoft System Center website and click SCOM 2007. Save AWS-MP-Setup-2.5.msi, and then run it.
2. Click Next and follow the directions to upgrade the components that you installed previously.
3. If your root management server, Operations console, and watcher node are on different computers, you must download and run the setup program on each computer.
4. On the watcher node, open a Command Prompt window as an administrator and run the following commands.

```
C:\> net stop HealthService
The System Center Management service is stopping.
The System Center Management service was stopped successfully.
C:\> net start HealthService
The System Center Management service is starting.
The System Center Management service was started successfully.
```
5. In the Operations console, on the Go menu, click Administration, and then click Management Packs.
6. In the Actions pane, click Import Management Packs.
7. On the Select Management Packs page, click Add, and then click Add from disk.
8. In the Select Management Packs to import dialog box, change the directory to C:\Program Files (x86)\Amazon Web Services Management Pack, select the Amazon.AmazonWebServices.mp file, and then click Open.
9. On the Select Management Packs page, under Import list, select the Amazon Web Services management pack, and then click Install.
   
   If the Install button is disabled, upgrading to the current version is not supported and you must uninstall AWS Management Pack first. For more information, see Uninstalling the AWS Management Pack (p. 1026).

Uninstalling the AWS Management Pack

If you need to uninstall the AWS Management Pack, use the following procedure.

System Center 2012

To uninstall the AWS Management Pack

1. In the Operations console, on the Go menu, click Administration, and then click Management Packs.
2. Right-click Amazon Web Services and select Delete.
3. In the Dependent Management Packs dialog box, note the dependent management packs, and then click Close.
4. Right-click the dependent management pack and select Delete.
5. Right-click Amazon Web Services and select Delete.
System Center 2007 R2

To uninstall the AWS Management Pack

1. Complete steps 1 through 5 described for System Center 2012 in the previous section.
2. From Control Panel, open Programs and Features. Select Amazon Web Services Management Pack and then click Uninstall.
3. If your root management server, Operations console, and watcher node are on different computers, you must repeat this process on each computer.

Troubleshooting the AWS Management Pack

The following are common errors, events, and troubleshooting steps.

Contents
- Errors 4101 and 4105 (p. 1027)
- Error 4513 (p. 1027)
- Event 623 (p. 1028)
- Events 2023 and 2120 (p. 1028)
- Event 6024 (p. 1028)
- General Troubleshooting for System Center 2012 — Operations Manager (p. 1028)
- General Troubleshooting for System Center 2007 R2 (p. 1029)

Errors 4101 and 4105

If you receive one of the following errors, you must upgrade the AWS Management Pack. For more information, see Upgrading the AWS Management Pack (p. 1025).

Error 4101
Exception calling "DescribeVolumes" with "1" argument(s): "AWS was not able to validate the provided access credentials"

Error 4105
Exception calling "DescribeApplications" with "0" argument(s): "The security token included in the request is invalid"

Error 4513

If you receive one of the following error, you must upgrade the AWS Management Pack. For more information, see Upgrading the AWS Management Pack (p. 1025).

Error 4513
The callback method DeliverDataToModule failed with exception "Resolution of the dependency failed, type = "Amazon.SCOM.SDK.Interfaces.IMonitorSdk", name = "(none)". Exception occurred while: Calling constructor Amazon.SCOM.SDK.CloudWatch.AwsMonitorSdk(System.String awsAccessKey, System.String awsSecretKey). Exception is: InvalidOperationException - Collection was modified; enumeration operation
Event 623

If you find the following event in the Windows event log, follow the solution described in KB975057.

Event ID: 623
HealthService (process_id) The version store for instance instance ("name") has reached its maximum size of size MB. It is likely that a long-running transaction is preventing cleanup of the version store and causing it to build up in size. Updates will be rejected until the long-running transaction has been completely committed or rolled back.
Possible long-running transaction:
SessionId: id
Session-context: value
Session-context ThreadId: id
Cleanup: value

Events 2023 and 2120

If you find the following events in the Windows event log, see Event ID 2023 and 2120 for more information.

Event ID: 2023
The Health Service has removed some items from the send queue for management group "Servers" since it exceeded the maximum allowed size of size megabytes.

Event ID: 2120
The Health Service has deleted one or more items for management group "Servers" which could not be sent in 1440 minutes.

Event 6024

If you find the following event in the Windows event log, see SCOM 2012 - Event ID 6024 for more information.

Event ID: 6024
LaunchRestartHealthService.js : Launching Restart Health Service. Health Service exceeded Process\Handle Count or Private Bytes threshold.

General Troubleshooting for System Center 2012 — Operations Manager

Try the following to resolve any issues.

- Verify that you have installed the latest Update Rollup for System Center 2012 — Operations Manager. The AWS Management Pack requires at least Update Rollup 1.
- Ensure that you have configured the AWS Management Pack after importing it by running the Add Monitoring Wizard. For more information, see Step 1: Installing the AWS Management Pack (p. 996).
- Verify that you have waited long enough for the AWS resources to be discovered (10–20 minutes).
- Verify that the management servers are configured properly.
• Management servers must have Internet connectivity.
• The action account for a management server must have local administrator privileges on the management server.
• The management server must have the .NET Framework 4.5. or later.
• Verify that the AWS Run As account is valid.
• The values for the access key ID and secret access key are correct.
• The access keys are active: In the AWS Management Console, click your name in the navigation bar and then click Security Credentials.
• The IAM user has at least read-only access permission. Note that read-only access allows the user actions that do not change the state of a resource, such as monitoring, but do not allow the user actions like launching or stopping an instance.
• If an Amazon CloudWatch metric shows as Not Monitored, check whether at least one Amazon CloudWatch alarm has been defined for that Amazon CloudWatch metric.
• For further troubleshooting, use the information in the event logs.
• Check the Operations Manager event log on the management server. For more information, see Events (p. 1022) for a list of the events that the AWS Management Pack writes to the Operations Manager event log.

General Troubleshooting for System Center 2007 R2

Try the following to resolve any issues.

• Ensure that you have configured the AWS Management Pack after importing it by running the Add Monitoring Wizard. For more information, see Step 1: Installing the AWS Management Pack (p. 996).
• Verify that you have waited long enough for the AWS resources to be discovered (10–20 minutes).
• Verify that the watcher node is configured properly.
  • The proxy agent is enabled. For more information, see Step 2: Configuring the Watcher Node (p. 998).
  • The watcher node has Internet connectivity.
  • The action account for the watcher node has local administrator privileges.
  • The watcher node must have the .NET Framework 3.5.1 or later.
• Verify that the watcher node is healthy and resolve all alerts. For more information, see Views (p. 1006).
• Verify that the AWS Run As account is valid.
• The values for the access key ID and secret access key are correct.
• The access keys are active: In the AWS Management Console, click your name in the navigation bar and then click Security Credentials.
• The IAM user has at least read-only access permission. Note that read-only access allows the user actions that do not change the state of a resource, such as monitoring, but do not allow the user actions like launching or stopping an instance.
• If an Amazon CloudWatch metric shows as Not Monitored, check whether at least one Amazon CloudWatch alarm has been defined for that Amazon CloudWatch metric.
• For further troubleshooting, use the information in the event logs.
• Check the Operations Manager event log on the management server as well as the watcher node. For more information, see Events (p. 1022) for a list of the events that the AWS Management Pack writes to the Operations Manager event log.
Using EC2Rescue for Windows Server

EC2Rescue for Windows Server is an easy-to-use tool that you run on an Amazon EC2 Windows Server instance to diagnose and troubleshoot possible problems. It is valuable for collecting log files and troubleshooting issues and also proactively searching for possible areas of concern. It can even examine Amazon EBS root volumes from other instances and collect relevant logs for troubleshooting Windows Server instances using that volume.

EC2Rescue for Windows Server has two different modules: a data collector module that collects data from all different sources, and an analyzer module that parses the data collected against a series of predefined rules to identify issues and provide suggestions.

The EC2Rescue for Windows Server tool only runs on Amazon EC2 instances running Windows Server 2008 R2 and later. When the tool starts, it checks whether it is running on an Amazon EC2 instance.

Note
If you are using a Linux instance, see EC2Rescue for Linux.

Contents
• Using EC2Rescue for Windows Server GUI (p. 1030)
• Using EC2Rescue for Windows Server with the Command Line (p. 1033)
• Using EC2Rescue for Windows Server with Systems Manager Run Command (p. 1038)

Using EC2Rescue for Windows Server GUI

EC2Rescue for Windows Server can perform the following analysis on an offline instance:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnose and Rescue</td>
<td>EC2Rescue for Windows Server can detect and address issues with the following service settings:</td>
</tr>
<tr>
<td></td>
<td>• System Time</td>
</tr>
<tr>
<td></td>
<td>• RealTimeisUniversal - Detects whether the RealTimeisUniversal registry key is enabled. If disabled, Windows system time drifts when the timezone is set to a value other than UTC.</td>
</tr>
<tr>
<td></td>
<td>• Windows Firewall</td>
</tr>
<tr>
<td></td>
<td>• Domain networks - Detects whether this Windows Firewall profile is enabled or disabled.</td>
</tr>
<tr>
<td></td>
<td>• Private networks - Detects whether this Windows Firewall profile is enabled or disabled.</td>
</tr>
</tbody>
</table>
### Using the GUI

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guest or public networks</strong></td>
<td>Detects whether this Windows Firewall profile is enabled or disabled.</td>
</tr>
<tr>
<td><strong>Remote Desktop</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Service Start</strong></td>
<td>Detects whether the Remote Desktop service is enabled.</td>
</tr>
<tr>
<td><strong>Remote Desktop Connections</strong></td>
<td>Detects whether this is enabled.</td>
</tr>
<tr>
<td><strong>TCP Port</strong></td>
<td>Detects which port the Remote Desktop service is listening on.</td>
</tr>
<tr>
<td><strong>EC2Config (Windows Server 2012 R2 and earlier)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Installation</strong></td>
<td>Detects which EC2Config version is installed.</td>
</tr>
<tr>
<td><strong>Service Start</strong></td>
<td>Detects whether the EC2Config service is enabled.</td>
</tr>
<tr>
<td><strong>Ec2SetPassword</strong></td>
<td>Generates a new administrator password.</td>
</tr>
<tr>
<td><strong>Ec2HandleUserData</strong></td>
<td>Allows you to execute a user data script on the next boot of the instance.</td>
</tr>
<tr>
<td><strong>EC2Launch (Windows Server 2016 and later)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Installation</strong></td>
<td>Detects which EC2Launch version is installed.</td>
</tr>
<tr>
<td><strong>Ec2SetPassword</strong></td>
<td>Generates a new administrator password.</td>
</tr>
<tr>
<td><strong>Network Interface</strong></td>
<td></td>
</tr>
<tr>
<td><strong>DHCP Service Startup</strong></td>
<td>Detects whether the DHCP service is enabled.</td>
</tr>
<tr>
<td><strong>Ethernet detail</strong></td>
<td>Displays information about the network driver version, if detected.</td>
</tr>
<tr>
<td><strong>DHCP on Ethernet</strong></td>
<td>Detects whether DHCP is enabled.</td>
</tr>
</tbody>
</table>

### Restore

Perform one of the following actions:

- **Last Known Good Configuration** - Attempts to boot the instance into the last known bootable state.
- **Restore registry from backup** - Restores the registry from `\Windows\System32\config\RegBack`.

### Capture Logs

Allows you to capture logs on the instance for analysis.
EC2Rescue for Windows Server can collect the following data from active and offline instances:

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Log</td>
<td>Collects application, system, and EC2Config event logs.</td>
</tr>
<tr>
<td>Memory Dump</td>
<td>Collects any memory dump files that exist on the instance.</td>
</tr>
<tr>
<td>EC2Config File</td>
<td>Collects log files generated by the EC2Config service.</td>
</tr>
<tr>
<td>EC2Launch File</td>
<td>Collects log files generated by the EC2Launch scripts.</td>
</tr>
<tr>
<td>SSM Agent File</td>
<td>Collects log files generated by SSM Agent.</td>
</tr>
<tr>
<td>Sysprep Log</td>
<td>Collects log files generated by the Windows System Preparation tool.</td>
</tr>
<tr>
<td>Driver SetupAPI Log</td>
<td>Collects Windows SetupAPI logs (setupapi.dev.log and setupapi.setup.log).</td>
</tr>
<tr>
<td>Registry</td>
<td>Collects SYSTEM and SOFTWARE hives.</td>
</tr>
<tr>
<td>System Information</td>
<td>Collects MSInfo32.</td>
</tr>
<tr>
<td>Boot Configuration</td>
<td>Collects HKEY_LOCAL_MACHINE\BCD00000000 hive.</td>
</tr>
<tr>
<td>Windows Update Log</td>
<td>Collects information about the updates that are installed on the instance.</td>
</tr>
</tbody>
</table>

**Note**

Windows Update logs are not captured on Windows Server 2016 and later instances.

---

**Video Walkthrough**

Brandon shows you how to use the Diagnose and Rescue feature of EC2Rescue for Windows Server:

*AWS Knowledge Center Videos: How do I use the Diagnose and Rescue feature of EC2Rescue?*

**Analyzing an Offline Instance**

The **Offline Instance** option is useful for debugging boot issues with Windows instances.

**To perform an action on an offline instance**

1. From a working Windows Server instance, download the EC2Rescue for Windows Server tool and extract the files.

   **Note**
   You can run the following PowerShell command to download EC2Rescue without changing your Internet Explorer Enhanced Security Configuration (ESC):
Collecting Data from an Active Instance

You can collect logs and other data from an active instance.

**To collect data from an active instance**

1. Connect to your Windows instance.
2. Download the EC2Rescue for Windows Server tool to your Windows instance and extract the files.

   **Note**
   
   You can run the following PowerShell command to download EC2Rescue without changing your Internet Explorer Enhanced Security Configuration (ESC):
   
   ```powershell
   PS C:\> Invoke-WebRequest https://s3.amazonaws.com/ec2rescue/windows/EC2Rescue_latest.zip -OutFile $env:USERPROFILE\Desktop\EC2Rescue_latest.zip
   ```
   
   This command will download the EC2Rescue .zip file to the desktop of the currently logged in user.

3. Open the EC2Rescue for Windows Server application and accept the license agreement.
4. Choose *Next*, *Current instance*, *Capture logs*.
5. Select the data items to collect and choose *Collect...*. Read the warning and choose *Yes* to continue.
6. Choose a file name and location for the ZIP file and choose *Save*.
7. After EC2Rescue for Windows Server completes, choose *Open Containing Folder* to view the ZIP file.
8. Choose *Finish*.

**Using EC2Rescue for Windows Server with the Command Line**

The EC2Rescue for Windows Server command line interface (CLI) allows you to run an EC2Rescue for Windows Server plugin (referred as an "action") programmatically.

The EC2Rescue for Windows Server tool has two execution modes:
**Collect Action**

EC2Rescue for Windows Server can perform the following actions on an Amazon EC2 Windows instance:

- **Collect Action (p. 1034)**
- **Rescue Action (p. 1036)**
- **Restore Action (p. 1037)**

**Collect Action**

EC2Rescue for Windows Server can collect the following data from active and offline instances. You can collect all logs, an entire log group, or an individual log within a group.

<table>
<thead>
<tr>
<th>Log Group</th>
<th>Available Logs</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td></td>
<td>Collects all available logs.</td>
</tr>
<tr>
<td>system-info</td>
<td>'MSInfo32 Output'</td>
<td>Collects MSInfo32.</td>
</tr>
</tbody>
</table>
| eventlog  | • 'Application'  
            • 'System'  
            • 'EC2ConfigService' | Collects application, system, and EC2Config event logs. |
| memory-dump | • 'Memory Dump File'  
              • 'Mini Dump Files' | Collects any memory dump files that exist on the instance. |
| ec2config | • 'Log Files'  
             • 'Configuration Files' | Collects log files generated by the EC2Config service. |
| ec2launch | • 'Logs'  
             • 'Config' | Collects log files generated by the EC2Launch scripts. |
| ssm-agent | 'Log Files' | Collects log files generated by SSM Agent. |
| sysprep  | 'Log Files' | Collects log files generated by the Windows System Preparation tool. |
| driver-setup | • 'SetupAPI Log Files'  
                • 'DPInst Log File'  
                • 'AWS PV Setup Log File' | Collects Windows SetupAPI logs (setupapi.dev.log and setupapi.setup.log). |
| registry | • 'SYSTEM'  
            • 'SOFTWARE' | Collects SYSTEM and SOFTWARE hives. |
The following are the available options:

- `/output:<outputFilePath>` - Required destination file path location to save collected log files in zip format.
- `/no-offline` - Optional attribute used in offline mode. Does not set the volume offline after completing the action.
- `/no-fix-signature` - Optional attribute used in offline mode. Does not fix a possible disk signature collision after completing the action.

Examples

The following are examples using the EC2Rescue for Windows Server CLI.

Online Mode Examples

Collect all available logs:

```
EC2RescueCmd /accepteula /online /collect:all /output:<outputFilePath>
```

Collect only a specific log group:

```
EC2RescueCmd /accepteula /online /collect:ec2config /output:<outputFilePath>
```

Collect individual logs within a log group:

```
EC2RescueCmd /accepteula /online /collect:'ec2config.Log Files,driver-setup.SetupAPI Log Files' /output:<outputFilePath>
```

Offline Mode Examples

Collect all available logs from an EBS volume. The volume is specified by the device_id value.
Rescue Action

EC2Rescue for Windows Server can detect and address issues with the following service settings:

<table>
<thead>
<tr>
<th>Service Group</th>
<th>Available Actions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td></td>
<td></td>
</tr>
<tr>
<td>system-time</td>
<td>'RealTimeIsUniversal'</td>
<td>System Time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>firewall</td>
<td>'Domain networks'</td>
<td>Windows Firewall</td>
</tr>
<tr>
<td></td>
<td>'Private networks'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'Guest or public networks'</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rdp</td>
<td>'Service Start'</td>
<td>Remote Desktop</td>
</tr>
<tr>
<td></td>
<td>'Remote Desktop Connections'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'TCP Port'</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ec2config</td>
<td>'Service Start'</td>
<td>EC2Config</td>
</tr>
<tr>
<td></td>
<td>'Ec2SetPassword'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'Ec2HandleUserData'</td>
<td></td>
</tr>
</tbody>
</table>
Service Group | Available Actions | Description
--- | --- | ---
e2launch | 'Reset Administrator Password' | Generates a new Windows administrator password.

network | 'DHCP Service Startup' | Generates a new Windows administrator password.

The following are the available options:

- **/level:<level>** - Optional attribute for the check level that the action should trigger. Allowed values are: information, warning, error, all. By default, it is set to error.
- **/check-only** - Optional attribute that generates a report but makes no modifications to the offline volume.
- **/no-offline** - Optional attribute that prevents the volume from being set offline after completing the action.
- **/no-fix-signature** - Optional attribute that does not fix a possible disk signature collision after completing the action.

Rescue Examples

The following are examples using the EC2Rescue for Windows Server CLI. The volume is specified using the device_id value.

Attempt to fix all identified issues on a volume:

```
EC2RescueCmd /accepteula /offline:xvd /rescue:all
```

Attempt to fix all issues within a service group on a volume:

```
EC2RescueCmd /accepteula /offline:xvd /rescue:firewall
```

Attempt to fix a specific item within a service group on a volume:

```
EC2RescueCmd /accepteula /offline:xvd /rescue:rdp.'Service Start'
```

Specify multiple issues to attempt to fix on a volume:

```
EC2RescueCmd /accepteula /offline:xvd /rescue:'system-time.RealTimeIsUniversal,ec2config.Service Start'
```

Restore Action

EC2Rescue for Windows Server can detect and address issues with the following service settings:
### Service Group | Available Actions | Description
--- | --- | ---
Restore Last Known Good Configuration | `lkgc` | Last Known Good Configuration - Attempts to boot the instance into the last known bootable state.

| Service Group | Available Actions | Description |
--- | --- | ---
Restore Windows registry from latest backup | `regback` | Restore registry from backup - Restores the registry from Windows\System32\config\RegBack.

The following are the available options:

- `/no-offline`—Optional attribute that prevents the volume from being set offline after completing the action.
- `/no-fix-signature`—Optional attribute that does not fix a possible disk signature collision after completing the action.

### Restore Examples

The following are examples using the EC2Rescue for Windows Server CLI. The volume is specified using the device_id value.

- **Restore last known good configuration on a volume:**

  ```
  EC2RescueCmd /accepteula /offline:xvdf /restore:lkgc
  ```

- **Restore the last Windows registry backup on a volume:**

  ```
  EC2RescueCmd /accepteula /offline:xvdf /restore:regback
  ```

## Using EC2Rescue for Windows Server with Systems Manager Run Command

AWS Support provides you with a Systems Manager Run Command document to interface with your Systems Manager-enabled instance to run EC2Rescue for Windows Server. The Run Command document is called `AWSSupport-RunEC2RescueForWindowsTool`.

This Systems Manager Run Command document performs the following tasks:

- Downloads and verifies EC2Rescue for Windows Server.
- Imports a PowerShell module to ease your interaction with the tool.
- Runs EC2RescueCmd with the provided command and parameters.

The Systems Manager Run Command document accepts three parameters:

- **Command**—The EC2Rescue for Windows Server action. The current allowed values are:
- **ResetAccess**—Resets the local Administrator password. The local Administrator password of the current instance will be reset and the randomly generated password will be securely stored in Parameter Store as `/EC2Rescue/Password/<INSTANCE_ID>`. If you select this action and provide no parameters, passwords are encrypted automatically with the default KMS key. Optionally, you can specify a KMS Key ID in Parameters to encrypt the password with your own key.

- **CollectLogs**—Runs EC2Rescue for Windows Server with the `/collect:all` action. If you select this action, **Parameters** must include an Amazon S3 bucket name to upload the logs to.

- **FixAll**—Runs EC2Rescue for Windows Server with the `/rescue:all` action. If you select this action, **Parameters** must include the block device name to rescue.

- **Parameters**—The PowerShell parameters to pass for the specified command.

**Note**

In order for the **ResetAccess** action to work, your Amazon EC2 instance needs to have the following policy attached in order to write the encrypted password to Parameter Store. Please wait a few minutes before attempting to reset the password of an instance after you have attached this policy to the related IAM role.

Using the default KMS key:

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ssm:PutParameter"
            ],
            "Resource": [
                "arn:aws:ssm:region:account_id:parameter/EC2Rescue/Passwords/<instanceid>"
            ]
        }
    ]
}
```

Using a custom KMS key:

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ssm:PutParameter"
            ],
            "Resource": [
                "arn:aws:ssm:region:account_id:parameter/EC2Rescue/Passwords/<instanceid>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Encrypt"
            ],
            "Resource": [
                "arn:aws:kms:region:account_id:key/<kmskeyid>"
            ]
        }
    ]
}
```
The following procedure describes how to view the JSON for this document in the Amazon EC2 console.

**To view the JSON for the Systems Manager Run Command document**

2. In the navigation pane, expand **Shared Services** and choose **Documents**.
3. In the search bar, set **Owner** as **Owned by Me or Amazon** and set the **Document name prefix** to **AWSSupport-RunEC2RescueForWindowsTool**.
4. Select the **AWSSupport-RunEC2RescueForWindowsTool** document, choose **Contents**, and then view the JSON.

**Examples**

Here are some examples on how to use the Systems Manager Run Command document to execute EC2Rescue for Windows Server, using the AWS CLI. For more information about sending commands with the AWS CLI, see the AWS CLI Command Reference.

**Attempt to Fix All Identified Issues on an Offline Root Volume**

Attempt to fix all identified issues on an offline root volume attached to an Amazon EC2 Windows instance:

```bash
aws ssm send-command --instance-ids "i-0cb2b964d3e14fd9f" --document-name "AWSSupport-RunEC2RescueForWindowsTool" --comment "EC2Rescue offline volume xvdf" --parameters "Command=FixAll, Parameters='xvdf'" --output text
```

**Collect Logs from the Current Amazon EC2 Windows Instance**

Collect all logs from the current online Amazon EC2 Windows instance and upload them to an Amazon S3 bucket:

```bash
aws ssm send-command --instance-ids "i-0cb2b964d3e14fd9f" --document-name "AWSSupport-RunEC2RescueForWindowsTool" --comment "EC2Rescue online log collection to S3" --parameters "Command=CollectLogs, Parameters='YOURS3BUCKETNAME'" --output text
```

**Collect Logs from an Offline Amazon EC2 Windows Instance Volume**

Collect all logs from an offline volume attached to an Amazon EC2 Windows instance and upload them to Amazon S3 with a presigned URL:

```bash
aws ssm send-command --instance-ids "i-0cb2b964d3e14fd9f" --document-name "AWSSupport-RunEC2RescueForWindowsTool" --comment "EC2Rescue offline log collection to S3" --parameters "Command=CollectLogs, Parameters="-Offline -BlockDeviceName xvdf -S3PreSignedUrl YOURS3PREIGNEDURL"" --output text
```

**Reset the Local Administrator Password**

The following examples show methods you can use to reset the local Administrator password. The output provides a link to Parameter Store, where you can find the randomly generated secure password you can then use to RDP to your Amazon EC2 Windows instance as the local Administrator.

Reset the local Administrator password of an online instance using the default KMS key alias/aws/ssm:
Reset the local Administrator password of an online instance using a KMS key:

```
aws ssm send-command --instance-ids "i-0cb2b964d3e14fd9f" --document-name "AWSSupport-RunEC2RescueForWindowsTool" --comment "EC2Rescue online password reset" --parameters "Command=ResetAccess" --output text
```

**Note**

In this example, the KMS key is a133dc3c-a2g4-4fc6-a873-6c0720104bf0.
Troubleshooting Windows Instances

The following procedures and tips can help you troubleshoot problems with your Amazon EC2 Windows instances.

Contents
- Troubleshooting Instance Launch Issues (p. 1042)
- Troubleshooting Connecting to Your Windows Instance (p. 1045)
- Troubleshoot an Unreachable Instance (p. 1051)
- Resetting a Lost or Expired Windows Administrator Password (p. 1058)
- Troubleshooting Stopping Your Instance (p. 1066)
- Troubleshooting Terminating (Shutting Down) Your Instance (p. 1067)
- Sending a Diagnostic Interrupt (Advanced Users Only) (p. 1068)
- Common Issues (p. 1069)
- Common Messages (p. 1073)

To get additional information for troubleshooting problems with your instance, use Using EC2Rescue for Windows Server (p. 1030). For information about troubleshooting issues with PV drivers, see Troubleshooting PV Drivers (p. 441).

Troubleshooting Instance Launch Issues

The following issues prevent you from launching an instance.

Launch Issues
- Instance Limit Exceeded (p. 1042)
- Insufficient Instance Capacity (p. 1043)
- Instance Terminates Immediately (p. 1043)
- High CPU Usage Shortly After Windows Starts (p. 1044)

Instance Limit Exceeded

Description

You get the InstanceLimitExceeded error when you try to launch a new instance or restart a stopped instance.

Cause

If you get an InstanceLimitExceeded error when you try to launch a new instance or restart a stopped instance, you have reached the limit on the number of instances that you can launch in a region. When you create your AWS account, we set default limits on the number of instances you can run on a per-region basis.

Solution

You can request an instance limit increase on a per-region basis. For more information, see Amazon EC2 Service Limits (p. 974).
Insufficient Instance Capacity

Description
You get the InsufficientInstanceCapacity error when you try to launch a new instance or restart a stopped instance.

Cause
If you get an InsufficientInstanceCapacity error when you try to launch an instance or restart a stopped instance, AWS does not currently have enough available On-Demand capacity to service your request.

Solution
To resolve the issue, try the following:

- Wait a few minutes and then submit your request again; capacity can shift frequently.
- Submit a new request with a reduced number of instances. For example, if you're making a single request to launch 15 instances, try making 3 requests for 5 instances, or 15 requests for 1 instance instead.
- If you're launching an instance, submit a new request without specifying an Availability Zone.
- If you're launching an instance, submit a new request using a different instance type (which you can resize at a later stage). For more information, see Changing the Instance Type (p. 188).
- If you are launching instances into a cluster placement group, you can get an insufficient capacity error. For more information, see Placement Group Rules and Limitations (p. 713).
- Try creating an On-Demand Capacity Reservation, which enables you to reserve Amazon EC2 capacity for any duration. For more information, see On-Demand Capacity Reservations (p. 317).
- Try purchasing Reserved Instances, which are a long-term capacity reservation. For more information, see Amazon EC2 Reserved Instances.

Instance Terminates Immediately

Description
Your instance goes from the pending state to the terminated state immediately after restarting it.

Cause
The following are a few reasons why an instance might immediately terminate:

- You've reached your EBS volume limit.
- An EBS snapshot is corrupt.
- The root EBS volume is encrypted and you do not have permissions to access the KMS key for decryption.
- The instance store-backed AMI that you used to launch the instance is missing a required part (an image.part.xx file).

Solution
You can use the Amazon EC2 console or AWS Command Line Interface to get the termination reason.
To get the termination reason using the Amazon EC2 console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances, and select the instance.
3. In the Description tab, note the reason next to the State transition reason label.

To get the termination reason using the AWS Command Line Interface

1. Use the describe-instances command and specify the instance ID.
   ```
   aws ec2 describe-instances --instance-id instance_id
   ```
2. Review the JSON response returned by the command and note the values in the StateReason response element.

   The following code block shows an example of a StateReason response element.
   ```
   "StateReason": {
     "Code": "Server.InternalError"
   },
   ```

To address the issue

Take one of the following actions depending on the termination reason you noted:

- If the reason is Client.VolumeLimitExceeded: Volume limit exceeded, you have reached your EBS volume limit. For more information, see Instance Volume Limits (p. 928). To submit a request to increase your Amazon EBS volume limit, complete the AWS Support Center Create Case form. For more information, see Amazon EC2 Service Limits (p. 974).

- If the reason is Client.InternalError: Client error on launch, that typically indicates that the root volume is encrypted and that you do not have permissions to access the KMS key for decryption. To get permissions to access the required KMS key, add the appropriate KMS permissions to your IAM user. For more information, see Using Key Policies in AWS KMS in the AWS Key Management Service Developer Guide.

High CPU Usage Shortly After Windows Starts

If Windows Update is set to Check for updates but let me choose whether to download and install them (the default instance setting) this check can consume anywhere from 50 - 99% of the CPU on the instance. If this CPU consumption causes problems for your applications, you can manually change Windows Update settings in Control Panel or you can use the following script in the Amazon EC2 user data field:

```
reg add "HKLM\SOFTWARE\Microsoft\Windows\CurrentVersion\WindowsUpdate\Auto Update" /v AUOptions /t REG_DWORD /d 3 /f net stop wuauserv net start wuauserv
```

When you execute this script specify a value for /d. The default value is 3. Possible values include the following:

- Never check for updates
- Check for updates but let me choose whether to download and install them
• Download updates but let me choose whether to install them
• Install updates automatically

To modify the user data for an Amazon EBS-backed instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances, and select the instance.
3. Choose Actions, select Instance State, and then choose Stop.
4. In the confirmation dialog box, select Yes, Stop. It can take a few minutes for the instance to stop.
5. With the instance still selected, select Actions, select Instance Settings, and then choose View/Change User Data. Note that you can't change the user data if the instance is running, but you can view it.
6. In the View/Change User Data dialog box, update the user data, and then choose Save.

After you modify the user data for your instance, you can execute it. For more information, see User Data Execution (p. 473).

Troubleshooting Connecting to Your Windows Instance

The following are possible problems you may have and error messages you may see while trying to connect to your Windows instance.

Contents
• Remote Desktop Can't Connect to the Remote Computer (p. 1045)
• Error Using macOS RDP Client (p. 1048)
• RDP Displays a Black Screen Instead of the Desktop (p. 1048)
• Unable to Remotely Log On to an Instance with a User Account That Is Not an Administrator (p. 1048)
• Troubleshooting Remote Desktop Issues Using AWS Systems Manager (p. 1048)

Remote Desktop Can't Connect to the Remote Computer

Try the following to resolve issues related to connecting to your instance:

• Verify that you're using the correct public DNS hostname. (In the Amazon EC2 console, select the instance and check Public DNS (IPv4) in the details pane.) If your instance is in a VPC and you do not see a public DNS name, you must enable DNS hostnames. For more information, see Using DNS with Your VPC in the Amazon VPC User Guide.
• Verify that your instance has a public IPv4 address. If not, you can associate an Elastic IP address with your instance. For more information, see Elastic IP Addresses (p. 674).
• To connect to your instance using an IPv6 address, check that your local computer has an IPv6 address and is configured to use IPv6. If you launched an instance from a Windows Server 2008 SP2 AMI or earlier, your instance is not automatically configured to recognize an IPv6 address assigned to the instance. For more information, see Configure IPv6 on Your Instances in the Amazon VPC User Guide.
Remote Desktop Can't Connect to the Remote Computer

- Verify that your security group has a rule that allows RDP access. For more information, see Create a Security Group (p. 17).

- If you copied the password but get the error Your credentials did not work, try typing them manually when prompted. It's possible that you missed a character or got an extra white space character when you copied the password.

- Verify that the instance has passed status checks. For more information, see Status Checks for Your Instances (p. 535) and Troubleshooting Instances with Failed Status Checks (Amazon EC2 User Guide for Linux Instances).

- Verify that the route table for the subnet has a route that sends all traffic destined outside the VPC to the internet gateway for the VPC. For more information, see Creating a Custom Route Table (Internet Gateways) in the Amazon VPC User Guide.

- Verify that Windows Firewall, or other firewall software, is not blocking RDP traffic to the instance. We recommend that you disable Windows Firewall and control access to your instance using security group rules. You can use AWSSupport-TroubleshootRDP (p. 1049) to disable the Windows Firewall profiles using SSM Agent. To disable Windows Firewall on a Windows instance that is not configured for AWS Systems Manager, use AWSSupport-ExecuteEC2Rescue (p. 1051), or use the following manual steps:

Manual Steps

1. Stop the affected instance and detach its root volume.
2. Launch a temporary instance in the same Availability Zone as the affected instance.

   **Warning**

   If your temporary instance is based on the same AMI that the original instance is based on, you must complete additional steps or you won't be able to boot the original instance after you restore its root volume because of a disk signature collision. Alternatively, select a different AMI for the temporary instance. For example, if the original instance uses the AWS Windows AMI for Windows Server 2008 R2, launch the temporary instance using the AWS Windows AMI for Windows Server 2012.

3. Attach the root volume from the affected instance to this temporary instance. Connect to the temporary instance, open the Disk Management utility, and bring the drive online.

4. Open Regedit and select HKEY_LOCAL_MACHINE. From the File menu, choose Load Hive. Select the drive, open the file Windows\System32\config\SYSTEM, and specify a key name when prompted (you can use any name).

5. Select the key you just loaded and navigate to ControlSet001\Services\SharedAccess \Parameters\FirewallPolicy. For each key with a name of the form xxxxProfile, select the key and change EnableFirewall from 1 to 0. Select the key again, and from the File menu, choose Unload Hive.

6. (Optional) If your temporary instance is based on the same AMI that the original instance is based on, you must complete the following steps or you won't be able to boot the original instance after you restore its root volume because of a disk signature collision.

   **Warning**

   The following procedure describes how to edit the Windows Registry using Registry Editor. If you are not familiar with the Windows Registry or how to safely make changes using Registry Editor, see Configure the Registry.

   a. Open a command prompt, type regedit.exe, and press Enter.
   b. In the Registry Editor, choose HKEY_LOCAL_MACHINE from the context menu (right-click), and then choose Find.
   c. Type Windows Boot Manager and then choose Find Next.
   d. Choose the key named 11000001. This key is a sibling of the key you found in the previous step.
e. In the right pane, choose Element and then choose Modify from the context menu (right-click).

f. Locate the four-byte disk signature at offset 0x38 in the data. Reverse the bytes to create the disk signature, and write it down. For example, the disk signature represented by the following data is E9EB3AA5:

```
... 0030 00 00 00 00 01 00 00 00
0038 A5 3A EB E9 00 00 00 00
0040 00 00 00 00 00 00 00 00
...```

g. In a Command Prompt window, run the following command to start Microsoft DiskPart.

```
diskpart```

h. Run the following DiskPart command to select the volume. (You can verify that the disk number is 1 using the Disk Management utility.)

```
DiskPart> select disk 1
Disk 1 is now the selected disk.
```

i. Run the following DiskPart command to get the disk signature.

```
DiskPart> uniqueld disk
Disk ID: 0C764FA8```

j. If the disk signature shown in the previous step doesn't match the disk signature from BCD that you wrote down earlier, use the following DiskPart command to change the disk signature so that it matches:

```
DiskPart> uniqueld disk id=E9EB3AA5```

7. Using the Disk Management utility, bring the drive offline.

**Note**
The drive is automatically offline if the temporary instance is running the same operating system as the affected instance, so you won’t need to bring it offline manually.

8. Detach the volume from the temporary instance. You can terminate the temporary instance if you have no further use for it.

9. Restore the root volume of the affected instance by attaching it as /dev/sda1.

10. Start the instance.

- Verify that Network Level Authentication is disabled on instances that are not part of an Active Directory domain (use AWSSupport-TroubleshootRDP (p. 1049) to disable NLA).
- Verify that the Remote Desktop Service (TermService) Startup Type is Automatic and the service is started (use AWSSupport-TroubleshootRDP (p. 1049) to enable and start the RDP service).
- Verify that you are connecting to the correct Remote Desktop Protocol port, which by default is 3389 (use AWSSupport-TroubleshootRDP (p. 1049) to read the current RDP port and change it back to 3389).
- Verify that Remote Desktop connections are allowed on your instance (use AWSSupport-TroubleshootRDP (p. 1049) to enable Remote Desktop connections).
- Verify that the password has not expired. If the password has expired, you can reset it. For more information, see Resetting a Lost or Expired Windows Administrator Password (p. 1058).
• If you attempt to connect using a user account that you created on the instance and receive the error:
  The user cannot connect to the server due to insufficient access privileges, verify that you granted the user the right to log on locally. For more information, see Grant a Member the Right to Log On Locally.
• If you attempt more than the maximum allowed concurrent RDP sessions, your session is terminated with the message Your Remote Desktop Services session has ended. Another user connected to the remote computer, so your connection was lost. By default, you are allowed two concurrent RDP sessions to your instance.

Error Using macOS RDP Client

If you are connecting to a Windows Server 2012 R2 instance using the Remote Desktop Connection client from the Microsoft website, you may get the following error:

Remote Desktop Connection cannot verify the identity of the computer that you want to connect to.

Download the Microsoft Remote Desktop app from the Mac App Store and use the app to connect to your instance.

RDP Displays a Black Screen Instead of the Desktop

Try the following to resolve this issue:

• Check the console output for additional information. To get the console output for your instance using the Amazon EC2 console, select the instance, choose Actions, select Instance Settings, and then choose Get System Log.
• Verify that you are running the latest version of your RDP client.
• Try the default settings for the RDP client. For more information, see Remote Session Environment.
• If you are using Remote Desktop Connection, try starting it with the /admin option as follows.

`mstsc /v:instance /admin`

• If the server is running a full-screen application, it might have stopped responding. Use Ctrl+Shift+Esc to start Windows Task Manager, and then close the application.
• If the server is over-utilized, it might have stopped responding. To monitor the instance using the Amazon EC2 console, select the instance and then select the Monitoring tab. If you need to change the instance type to a larger size, see Changing the Instance Type (p. 188).

Unable to Remotely Log On to an Instance with a User Account That Is Not an Administrator

If you are not able to remotely log on to a Windows instance from a user account that is not an administrator account, ensure that you have granted the user the right to log on locally. See Grant a user or group the right to log on locally to the domain controllers in the domain.

Troubleshooting Remote Desktop Issues Using AWS Systems Manager

You can use AWS Systems Manager to troubleshoot issues connecting to your Windows instance using RDP.
AWSSupport-TroubleshootRDP

The AWSSupport-TroubleshootRDP automation document allows the user to check or modify common settings on the target instance that can impact Remote Desktop Protocol (RDP) connections, such as the RDP Port, Network Layer Authentication (NLA), and Windows Firewall profiles. By default, the document reads and outputs the values of these settings.

The AWSSupport-TroubleshootRDP automation document can be used only with instances that are enabled for use with AWS Systems Manager. These instances also referred to as managed instances. For more information, see Setting Up AWS Systems Manager for Hybrid Environments in the AWS Systems Manager User Guide. Your instance must have the AWS Systems Manager SSM Agent installed and have an IAM role attached with permissions to Systems Manager. You must change the default parameter values to modify the common RDP settings on the target instance.

To troubleshoot using the AWSSupport-TroubleshootRDP document

1. Log in to the Systems Manager Console.
2. Verify that you are in the same Region as the impaired instance.
3. Open the AWSSupport-TroubleshootRDP document.
4. In Execution Mode, choose Execute the entire automation at once.
5. In Input parameters, in the InstanceId field, enable Show interactive instance picker.
6. Choose your Amazon EC2 instance.
   
   **Note**
   If you don't see your instance in the list, it's not enabled for Systems Manager. For information, see Create an Instance Profile for Systems Manager and Attach an IAM Instance Profile to an Amazon EC2 Instance in the AWS Systems Manager User Guide.
7. Review the examples (p. 1049), then choose Execute automation.
8. To monitor the execution progress, in Execution status, wait for the status to change from Pending to Success. Expand Outputs to view the results. To view the output of individual steps, in Executed Steps, choose the Step ID.

AWSSupport-TroubleshootRDP Examples

The following examples show you how to accomplish common troubleshooting tasks using AWSSupport-TroubleshootRDP. You can use either the example AWS CLI start-automation-execution command or the provided link to the AWS Management Console.

Example Example: Check the current RDP status

AWS CLI:

```
aws ssm start-automation-execution --document-name "AWSSupport-TroubleshootRDP" --parameters "InstanceId=instance_id" --region region_code
```

AWS Systems Manager console:

```
```

Example Example: Disable the Windows Firewall

AWS CLI:


aws ssm start-automation-execution --document-name "AWSSupport-TroubleshootRDP" --parameters "InstanceId=instance_id,Firewall=Disable" --region region_code

AWS Systems Manager console:


Example Example: Disable Network Level Authentication

AWS CLI:

aws ssm start-automation-execution --document-name "AWSSupport-TroubleshootRDP" --parameters "InstanceId=instance_id,NLAModifyAction"


Example Example: Set RDP Service Startup Type to Automatic and start the RDP service

AWS CLI:

aws ssm start-automation-execution --document-name "AWSSupport-TroubleshootRDP" --parameters "InstanceId=instance_id,RDPServiceStartupType=Auto, RDPServiceAction=Start" --region region_code

AWS Systems Manager console:


Example Example: Restore the default RDP Port (3389)

AWS CLI:

aws ssm start-automation-execution --document-name "AWSSupport-TroubleshootRDP" --parameters "InstanceId=instance_id,RDPPortAction=Modify" --region region_code

AWS Systems Manager console:


Example Example: Allow remote connections

AWS CLI:

aws ssm start-automation-execution --document-name "AWSSupport-TroubleshootRDP" --parameters "InstanceId=instance_id,RemoteConnections=Enable" --region region_code

AWS Systems Manager console:
Troubleshoot an Unreachable Instance

If you are unable to reach your instance through SSH or RDP, you can capture a screenshot of your instance and view it as an image. This provides visibility as to the status of the instance, and allows for quicker troubleshooting.

- How to Take a Screenshot of an Unreachable Instance (p. 1051)
- Common Screenshots (p. 1052)

How to Take a Screenshot of an Unreachable Instance

If you are unable to reach your instance via SSH or RDP, you can capture a screenshot of your instance and view it as an image. The image can provide visibility as to the status of the instance, and allows for quicker troubleshooting. You can generate screenshots while the instance is running or after it has crashed. There is no data transfer cost for this screenshot. The image is generated in JPG format and is no larger than 100 kb. This feature is not supported when the instance is using an NVIDIA GRID driver. This feature is available in the following Regions:

- US East (N. Virginia) Region
- US East (Ohio) Region
- US West (Oregon) Region
To get a screenshot of a running instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the left navigation pane, choose Instances.
3. Select the instance to capture.
5. Choose Get Instance Screenshot. Right-click the image to download and save it.

To get a screenshot of a running instance using the command line

You can use one of the following commands. The returned output is base64-encoded. For more information about these command line interfaces, see Accessing Amazon EC2 (p. 3).

- get-console-screenshot (AWS CLI)
- GetConsoleScreenshot (Amazon EC2 Query API)

For API calls, the returned content is base64-encoded. For command line tools, the decoding is performed for you.

Common Screenshots

You can use the following information to help you troubleshoot an unreachable instance based on screenshots returned by the service.

- Log On Screen (Ctrl+Alt+Delete) (p. 1052)
- Recovery Console Screen (p. 1055)
- Windows Boot Manager Screen (p. 1056)
- Sysprep Screen (p. 1056)
- Getting Ready Screen (p. 1057)
- Windows Update Screen (p. 1058)
- Chkdsk (p. 1058)

Log On Screen (Ctrl+Alt+Delete)

Console Screenshot Service returned the following.
If an instance becomes unreachable during logon, there could be a problem with your network configuration or Windows Remote Desktop Services. An instance can also be unresponsive if a process is using large amounts of CPU.

**Network Configuration**

Use the following information, to verify that your AWS, Microsoft Windows, and local (or on-premises) network configurations aren’t blocking access to the instance.

### AWS Network Configuration

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Verify</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security group configuration</td>
<td>Verify that port 3389 is open for your security group. Verify you are connecting to the right public IP address. If the instance was not associated with an Elastic IP, the public IP changes after the instance stops/starts. For more information, see Remote Desktop Can’t Connect to the Remote Computer (p. 1045).</td>
</tr>
<tr>
<td>VPC configuration (Network ACLs)</td>
<td>Verify that the access control list (ACL) for your Amazon VPC is not blocking access. For information, see Network ACLs in the Amazon VPC User Guide.</td>
</tr>
<tr>
<td>VPN configuration</td>
<td>If you are connecting to your VPC using a virtual private network (VPN), verify VPN tunnel connectivity. For more information, see How do I troubleshoot VPN tunnel connectivity to an Amazon VPC?</td>
</tr>
</tbody>
</table>

### Windows Network Configuration

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Verify</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows Firewall</td>
<td>Verify that Windows Firewall isn't blocking connections to your instance. Disable Windows</td>
</tr>
<tr>
<td>Configuration</td>
<td>Verify</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Firewall as described in bullet 7 of the remote</td>
<td>See &quot;Remote Desktop Can't Connect to the Remote Computer (p. 1045).</td>
</tr>
<tr>
<td>desktop troubleshooting section,</td>
<td></td>
</tr>
<tr>
<td>Advanced TCP/IP configuration (Use of static IP)</td>
<td>The instance may be unresponsive because you configured a static IP address. For a VPC, Create a network interface (p. 691) and attach it to the instance (p. 693). For EC2 Classic, enable DHCP.</td>
</tr>
<tr>
<td>Local or On-Premises Network Configuration</td>
<td></td>
</tr>
<tr>
<td>Verify that a local network configuration isn't</td>
<td>Try to connect to another instance in the same VPC as your unreachable instance. If you can't access another instance, work with your local network administrator to determine whether a local policy is restricting access.</td>
</tr>
<tr>
<td>blocking access. Try to connect to another instance in</td>
<td></td>
</tr>
<tr>
<td>the same VPC as your unreachable instance. If you can't</td>
<td></td>
</tr>
<tr>
<td>access another instance, work with your local network</td>
<td></td>
</tr>
<tr>
<td>administrator to determine whether a local policy is</td>
<td></td>
</tr>
<tr>
<td>restricting access.</td>
<td></td>
</tr>
<tr>
<td>Remote Desktop Services Issue</td>
<td></td>
</tr>
<tr>
<td>If the instance can't be reached during logon, there</td>
<td></td>
</tr>
<tr>
<td>could a problem with Remote Desktop Services (RDS) on</td>
<td></td>
</tr>
<tr>
<td>the instance.</td>
<td></td>
</tr>
<tr>
<td>Remote Desktop Services Configuration</td>
<td></td>
</tr>
<tr>
<td>RDS is running</td>
<td>Verify that RDS is running on the instance. Connect to the instance using the Microsoft Management Console (MMC) Services snap-in (services.msc). In the list of services, verify that Remote Desktop Services is Running. If it isn't, start it and then set the startup type to Automatic. If you can't connect to the instance by using the Services snap-in, detach the root volume from the instance, take a snapshot of the volume or create an AMI from it, attach the original volume to another instance in the same Availability Zone as a secondary volume, and modify the Start registry key. When you are finished, reattach the root volume to the original instance. For more information about detaching volumes, see Detaching an Amazon EBS Volume from an Instance (p. 816).</td>
</tr>
<tr>
<td>RDS is enabled</td>
<td>Even if the service is started, it might be disabled. Detach the root volume from the instance, take a snapshot of the volume or create an AMI from it, attach the original volume to another instance in the same Availability Zone as a secondary volume, and enable the service by modifying the Terminal Server registry key as described in the following article: Enable Remote Desktop Remotely By Using Remote Registry. When you are finished, reattach the root volume to the original instance. For more information, see Detaching an Amazon EBS Volume from an Instance (p. 816).</td>
</tr>
<tr>
<td>High CPU</td>
<td></td>
</tr>
<tr>
<td>Check the CPUUtilization (Maximum) metric on your</td>
<td>If CPUUtilization (Maximum) is a high number, wait for the CPU to go down and try connecting again. High CPU usage can be caused by:</td>
</tr>
<tr>
<td>instance by using Amazon CloudWatch. If</td>
<td></td>
</tr>
<tr>
<td>High CPU usage can be caused by:</td>
<td></td>
</tr>
</tbody>
</table>
Amazon Elastic Compute Cloud
User Guide for Windows Instances
Common Screenshots

- Windows Update
- Security Software Scan
- Custom Startup Script
- Task Scheduler

For more information, see Get Statistics for a Specific Resource in the Amazon CloudWatch User Guide. For additional troubleshooting tips, see High CPU Usage Shortly After Windows Starts (p. 1044).

Recovery Console Screen

Console Screenshot Service returned the following.

![System Recovery Options](image)

![Choose an option](image)

The operating system may boot into the Recovery console and get stuck in this state if the bootstatuspolicy is not set to ignoreallfailures. Use the following procedure to change the bootstatuspolicy configuration to ignoreallfailures.

By default, the policy configuration for AWS-provided public Windows AMIs is set to ignoreallfailures.

1. Stop the unreachable instance.
2. Create a snapshot of the root volume. The root volume is attached to the instance as /dev/sda1.
   
   Detach the root volume from the unreachable instance, take a snapshot of the volume or create an AMI from it, and attach it to another instance in the same Availability Zone as a secondary volume. For more information, see Detaching an Amazon EBS Volume from an Instance (p. 816).

   **Warning**
   
   If your temporary instance is based on the same AMI that the original instance is based on, you must complete additional steps or you won't be able to boot the original instance after you restore its root volume because of a disk signature collision. Alternatively, select a different AMI for the temporary instance. For example, if the original instance uses an AMI for Windows Server 2008 R2, launch the temporary instance using an AMI for Windows Server 2012. If you must create a temporary instance based on the same AMI, see Step 6 in Remote Desktop Can't Connect to the Remote Computer (p. 1045) to avoid a disk signature collision.

3. Log in to the instance and execute the following command from a command prompt to change the bootstatuspolicy configuration to ignoreallfailures:
4. Reattach the volume to the unreachable instance and start the instance again.

## Windows Boot Manager Screen

Console Screenshot Service returned the following.

---

**Windows Boot Manager**

Windows failed to start. A recent hardware or software change might be the cause. To fix the problem:

1. Insert your Windows installation disc and restart your computer.
2. Choose your language settings, and then click "Next."
3. Click "Repair your computer."

If you do not have this disc, contact your system administrator or computer manufacturer for assistance.

File: `\Boot\BCD`
Status: 0xc000000f
Info: The Boot Configuration Data for your PC is missing or contains errors.

---

The operating system experienced a fatal corruption in the system file and/or the registry. When the instance is stuck in this state, you should recover the instance from a recent backup AMI or launch a replacement instance. If you need to access data on the instance, detach any root volumes from the unreachable instance, take a snapshot of those volume or create an AMI from them, and attach them to another instance in the same Availability Zone as a secondary volume. For more information, see [Detaching an Amazon EBS Volume from an Instance](p. 816).

## Sysprep Screen

Console Screenshot Service returned the following.

---

```bash
bcdedit /store Drive Letter:\boot\bcd /set {default} bootstatuspolicy ignoreallfailures
```
You may see this screen if you did not use the EC2Config Service to call sysprep.exe or if the operating system failed while running Sysprep. To solve this problem, Create a Standard Amazon Machine Image Using Sysprep (p. 113).

**Getting Ready Screen**

Console Screenshot Service returned the following.

Refresh the Instance Console Screenshot Service repeatedly to verify that the progress ring is spinning. If the ring is spinning, wait for the operating system to start up. You can also check the **CPUUtilization (Maximum)** metric on your instance by using Amazon CloudWatch to see if the operating system is active. If the progress ring is not spinning, the instance may be stuck at the boot process. Reboot the instance. If rebooting does not solve the problem, recover the instance from a recent backup AMI or launch a replacement instance. If you need to access data on the instance, detach the root volume from the unreachable instance, take a snapshot of the volume or create an AMI from it. Then attach it to another instance in the same Availability Zone as a secondary volume.
Windows Update Screen

Console Screenshot Service returned the following.

The Windows Update process is updating the registry. Wait for the update to finish. Do not reboot or stop the instance as this may cause data corruption during the update.

**Note**
The Windows Update process can consume resources on the server during the update. If you experience this problem often, consider using faster instance types and faster EBS volumes.

Chkdsk

Console Screenshot Service returned the following.

Windows is running the chkdsk system tool on the drive to verify file system integrity and fix logical file system errors. Wait for process to complete.

Resetting a Lost or Expired Windows Administrator Password

If you are no longer able to access your Windows Amazon EC2 instance because the Windows administrator password is lost or expired, you can reset the password.

**Note**
There is an AWS Systems Manager Automation document that automatically applies the manual steps necessary to reset the local administrator password. For more information, see Reset Passwords and SSH Keys on Amazon EC2 Instances in the AWS Systems Manager User Guide.

The manual methods to reset the administrator password use either EC2Config or EC2Launch.

- For Windows AMIs before Windows Server 2016, use the EC2Config service.
- For Windows Server 2016 and later AMIs, use the EC2Launch service.

These procedures also describe how to connect to an instance if you lost the key pair that was used to create the instance. Amazon EC2 uses a public key to encrypt a piece of data, such as a password, and a private key to decrypt the data. The public and private keys are known as a *key pair*. With Windows instances, you use a key pair to obtain the administrator password and then log in using RDP.
Resetting the Windows Administrator Password Using EC2Config

If you have lost your Windows administrator password and are using a Windows AMI before Windows Server 2016, you can use the EC2Config service to generate a new password.

If you are using a Windows Server 2016 or later AMI, see Resetting the Windows Administrator Password Using EC2Launch (p. 1063).

To reset your Windows administrator password using EC2Config, you need to do the following:

- Step 1: Verify That the EC2Config Service Is Running (p. 1059)
- Step 2: Detach the Root Volume from the Instance (p. 1060)
- Step 3: Attach the Volume to a Temporary Instance (p. 1060)
- Step 4: Modify the Configuration File (p. 1061)
- Step 5: Restart the Original Instance (p. 1062)

Step 1: Verify That the EC2Config Service Is Running

Before you attempt to reset the administrator password, verify that the EC2Config service is installed and running. You use the EC2Config service to reset the administrator password later in this section.

To verify that the EC2Config service is running

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances and then choose the instance that needs a password reset. This instance is referred to as the original instance in this procedure.
4. Locate the EC2 Agent entry, for example, EC2 Agent: Ec2Config service v3.18.1118. If you see this entry, the EC2Config service is running.
If the system log output is empty, or if the EC2Config service is not running, troubleshoot the instance using the Instance Console Screenshot service. For more information, see Troubleshoot an Unreachable Instance (p. 1051).

### Step 2: Detach the Root Volume from the Instance

You can't use EC2Config to reset an administrator password if the volume on which the password is stored is attached to an instance as the root volume. You must detach the volume from the original instance before you can attach it to a temporary instance as a secondary volume.

#### To detach the root volume from the instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose **Instances**.
3. Select the instance that needs a password reset and choose **Actions, Instance State, Stop**. After the instance's status changes to **stopped**, continue with the next step.
4. (Optional) If you have the private key that you specified when you launched this instance, continue with the next step. Otherwise, use the following steps to replace the instance with a new instance that you launch with a new key pair.
   a. Create a new key pair using the Amazon EC2 console. To give your new key pair the same name as the one for which you lost the private key, you must first delete the existing key pair.
   b. Select the instance to replace. Note the instance type, VPC, subnet, security group, and IAM role of the instance.
   c. Choose **Actions, Image, Create Image**. Type a name and a description for the image and choose **Create Image, View pending image**. After the image status changes to **available**, continue to the next step.
   d. Select the image and choose **Launch**.
   e. Complete the wizard, selecting the same instance type, VPC, subnet, security group, and IAM role as the instance to replace, and then choose **Launch**.
   f. When prompted, choose the key pair that you created for the new instance, select the acknowledgement check box, and then choose **Launch Instances**.
   g. (Optional) If the original instance has an associated Elastic IP address, transfer it to the new instance. If the original instance has EBS volumes in addition to the root volume, transfer them to the new instance.
   h. Terminate the stopped instance, as it is no longer needed. For the remainder of this procedure, all references to the original instance apply to this instance that you just created.
5. Detach the root volume from the original instance as follows:
   a. In the **Description** pane of the original instance, note the ID of the EBS volume listed as the **Root device**.
   b. In the navigation pane, choose **Volumes**.
   c. In the list of volumes, select the volume noted in the previous step, and choose **Actions, Detach Volume**. After the volume status changes to **available**, continue with the next step.

### Step 3: Attach the Volume to a Temporary Instance

Next, launch a temporary instance and attach the volume to it as a secondary volume. This is the instance you use to modify the configuration file.
To launch a temporary instance and attach the volume

1. Launch the temporary instance as follows:
   a. In the navigation pane, choose Instances, choose Launch Instance, and then select an AMI.
      Important
      To avoid disk signature collisions, you must select an AMI for a different version of Windows. For example, if the original instance runs Windows Server 2012 R2, launch the temporary instance using the base AMI for Windows Server 2008 R2.
   b. Leave the default instance type and choose Next: Configure Instance Details.
   c. On the Configure Instance Details page, for Subnet, select the same Availability Zone as the original instance and choose Review and Launch.
      Important
      The temporary instance must be in the same Availability Zone as the original instance. If your temporary instance is in a different Availability Zone, you can't attach the original instance's root volume to it.
   e. When prompted, create a new key pair, download it to a safe location on your computer, and then choose Launch Instances.

2. Attach the volume to the temporary instance as a secondary volume as follows:
   a. In the navigation pane, choose Volumes, select the volume that you detached from the original instance, and then choose Actions, Attach Volume.
   b. In the Attach Volume dialog box, for Instances, start typing the name or ID of your temporary instance and select the instance from the list.
   c. For Device, type xvdf (if it isn't already there), and choose Attach.

Step 4: Modify the Configuration File

After you have attached the volume to the temporary instance as a secondary volume, modify the Ec2SetPassword plugin in the configuration file.

To modify the configuration file

1. From the temporary instance, modify the configuration file on the secondary volume as follows:
   a. Launch and connect to the temporary instance.
   b. Open the Disk Management utility, and bring the drive online using these instructions: Making an Amazon EBS Volume Available for Use.
   c. Navigate to the secondary volume, and open \Program Files\Amazon\Ec2ConfigService\Settings\config.xml using a text editor, such as Notepad.
   d. At the top of the file, find the plugin with the name Ec2SetPassword, as shown in the screenshot. Change the state from Disabled to Enabled and save the file.
2. After you have modified the configuration file, detach the secondary volume from the temporary instance as follows:

   a. Using the **Disk Management** utility, bring the volume offline.
   
   b. Disconnect from the temporary instance and return to the Amazon EC2 console.
   
   c. In the navigation pane, choose **Volumes**, select the volume, and then choose **Actions, Detach Volume**. After the volume's status changes to **available**, continue with the next step.

**Step 5: Restart the Original Instance**

After you have modified the configuration file, reattach the volume to the original instance as the root volume and connect to the instance using its key pair to retrieve the administrator password.

1. Reattach the volume to the original instance as follows:

   a. In the navigation pane, choose **Volumes**, select the volume that you detached from the temporary instance, and then choose **Actions, Attach Volume**.
   
   b. In the **Attach Volume** dialog box, for **Instances**, start typing the name or ID of your original instance and then select the instance.
   
   c. For **Device**, type `/dev/sda1`.
   
   d. Choose **Attach**. After the volume status changes to **in-use**, continue to the next step.

2. In the navigation pane, choose **Instances**. Select the original instance and choose **Actions, Instance State, Start**. When prompted for confirmation, choose **Yes, Start**. After the instance state changes to **running**, continue to the next step.

3. Retrieve your new Windows administrator password using the private key for the new key pair and connect to the instance. For more information, see Connecting to Your Windows Instance (p. 379).

   **Important**
   
   The instance gets a new public IP address after you stop and start it. Make sure to connect to the instance using its current public DNS name. For more information, see Instance Lifecycle (p. 329).

4. (Optional) If you have no further use for the temporary instance, you can terminate it. Select the temporary instance, and choose **Actions, Instance State, Terminate**.
Resetting the Windows Administrator Password Using EC2Launch

If you have lost your Windows administrator password and are using a Windows Server 2016 or later AMI, you can use the EC2Rescue tool, which uses the EC2Launch service to generate a new password.

If you are using a Windows Server AMI earlier than Windows Server 2016, see Resetting the Windows Administrator Password Using EC2Config (p. 1059).

Warning
When you stop an instance, the data on any instance store volumes is erased. To keep data from instance store volumes, be sure to back it up to persistent storage.

Note
If you have disabled the local administrator account on the instance and your instance is configured for Systems Manager, you can also re-enable and reset your local administrator password by using EC2Rescue and Run Command. For more information, see Using EC2Rescue for Windows Server with Systems Manager Run Command.

Note
There is an AWS Systems Manager Automation document that automatically applies the manual steps necessary to reset the local administrator password. For more information, see Reset Passwords and SSH Keys on Amazon EC2 Instances in the AWS Systems Manager User Guide.

To reset your Windows administrator password using EC2Launch, you need to do the following:

- Step 1: Detach the Root Volume from the Instance (p. 1063)
- Step 2: Attach the Volume to a Temporary Instance (p. 1064)
- Step 3: Reset the Administrator Password (p. 1065)
- Step 4: Restart the Original Instance (p. 1065)

Step 1: Detach the Root Volume from the Instance

You can't use EC2Launch to reset an administrator password if the volume on which the password is stored is attached to an instance as the root volume. You must detach the volume from the original instance before you can attach it to a temporary instance as a secondary volume.

To detach the root volume from the instance

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Select the instance that needs a password reset and choose Actions, Instance State, Stop. After the instance's status changes to stopped, continue with the next step.
4. (Optional) If you have the private key that you specified when you launched this instance, continue with the next step. Otherwise, use the following steps to replace the instance with a new instance that you launch with a new key pair.
   a. Create a new key pair using the Amazon EC2 console. To give your new key pair the same name as the one for which you lost the private key, you must first delete the existing key pair.
   b. Select the instance to replace. Note the instance type, VPC, subnet, security group, and IAM role of the instance.
   c. Choose Actions, Image, Create Image. Type a name and a description for the image and choose Create Image, View pending image. After the image status changes to available, continue to the next step.
d. Select the image and choose **Launch**.

e. Complete the wizard, selecting the same instance type, VPC, subnet, security group, and IAM role as the instance to replace, and then choose **Launch**.

f. When prompted, choose the key pair that you created for the new instance, select the acknowledgement check box, and then choose **Launch Instances**.

g. (Optional) If the original instance has an associated Elastic IP address, transfer it to the new instance. If the original instance has EBS volumes in addition to the root volume, transfer them to the new instance.

h. Terminate the stopped instance, as it is no longer needed. For the remainder of this procedure, all references to the original instance apply to this instance that you just created.

5. Detach the root volume from the original instance as follows:

a. In the **Description** pane of the original instance, note the ID of the EBS volume listed as the **Root device**.

b. In the navigation pane, choose **Volumes**.

c. In the list of volumes, select the volume noted in the previous step, and choose **Actions, Detach Volume**. After the volume status changes to **available**, continue with the next step.

**Step 2: Attach the Volume to a Temporary Instance**

Next, launch a temporary instance and attach the volume to it as a secondary volume. This is the instance you use to run EC2Launch.

**To launch a temporary instance and attach the volume**

1. Launch the temporary instance as follows:

   a. In the navigation pane, choose **Instances**, choose **Launch Instance**, and then select an AMI.

      **Important**
      To avoid disk signature collisions, you must select an AMI for a different version of Windows. For example, if the original instance runs Windows Server 2012 R2, launch the temporary instance using the base AMI for Windows Server 2008 R2.

   b. Leave the default instance type and choose **Next: Configure Instance Details**.

   c. On the **Configure Instance Details** page, for **Subnet**, select the same Availability Zone as the original instance and choose **Review and Launch**.

      **Important**
      The temporary instance must be in the same Availability Zone as the original instance. If your temporary instance is in a different Availability Zone, you can't attach the original instance's root volume to it.

   d. On the **Review Instance Launch** page, choose **Launch**.

   e. When prompted, create a new key pair, download it to a safe location on your computer, and then choose **Launch Instances**.

2. Attach the volume to the temporary instance as a secondary volume as follows:

   a. In the navigation pane, choose **Volumes**, select the volume that you detached from the original instance, and then choose **Actions, Attach Volume**.

   b. In the **Attach Volume** dialog box, for **Instances**, start typing the name or ID of your temporary instance and select the instance from the list.

   c. For **Device**, type `xvdf` (if it isn't already there), and choose **Attach**.
Step 3: Reset the Administrator Password

Next, connect to the temporary instance and use EC2Launch to reset the administrator password.

To reset the administrator password

1. Connect to the temporary instance and use the EC2Rescue for Windows Server tool on the instance to reset the administrator password as follows:
   a. Download the EC2Rescue for Windows Server zip file, extract the contents, and run EC2Rescue.exe.
   b. On the License Agreement screen, read the license agreement, and, if you accept the terms, choose I Agree.
   c. On the Welcome to EC2Rescue for Windows Server screen, choose Next.
   d. On the Select mode screen, choose Offline instance.
   e. On the Select a disk screen, select the xvdf device and choose Next.
   f. Confirm the disk selection and choose Yes.
   g. After the volume has loaded, choose OK.
   h. On the Select Offline Instance Option screen, choose Diagnose and Rescue.
   i. On the Summary screen, review the information and choose Next.
   j. On the Detected possible issues screen, select Reset Administrator Password and choose Next.
   k. On the Confirm screen, choose Rescue, OK.
   l. On the Done screen, choose Finish.
   m. Close the EC2Rescue for Windows Server tool, disconnect from the temporary instance, and then return to the Amazon EC2 console.

2. Detach the secondary (xvdf) volume from the temporary instance as follows:
   a. In the navigation pane, choose Instances and select the temporary instance.
   b. In the Description pane of the temporary instance, note the ID of the EBS volume listed as xvdf.
   c. In the navigation pane, choose Volumes.
   d. In the list of volumes, select the volume noted in the previous step, and choose Actions, Detach Volume. After the volume status changes to available, continue with the next step.

Step 4: Restart the Original Instance

After you have reset the administrator password using EC2Launch, reattach the volume to the original instance as the root volume and connect to the instance using its key pair to retrieve the administrator password.

To restart the original instance

1. Reattach the volume to the original instance as follows:
   a. In the navigation pane, choose Volumes, select the volume that you detached from the temporary instance, and then choose Actions, Attach Volume.
   b. In the Attach Volume dialog box, for Instances, start typing the name or ID of your original instance and then select the instance.
   c. For Device, type /dev/sda1.
   d. Choose Attach. After the volume status changes to in-use, continue to the next step.

2. In the navigation pane, choose Instances. Select the original instance and choose Actions, Instance State, Start. When prompted for confirmation, choose Yes, Start. After the instance state changes to running, continue to the next step.
3. Retrieve your new Windows administrator password using the private key for the new key pair and connect to the instance. For more information, see Connecting to Your Windows Instance (p. 379).

4. (Optional) If you have no further use for the temporary instance, you can terminate it. Select the temporary instance, and choose Actions, Instance State, Terminate.

Troubleshooting Stopping Your Instance

If you have stopped your Amazon EBS-backed instance and it appears stuck in the stopping state, there may be an issue with the underlying host computer.

There is no cost for any instance usage while an instance is not in the running state.

Force the instance to stop using either the console or the AWS CLI.

- To force the instance to stop using the console, select the stuck instance, and choose Actions, Instance State, Stop, and Yes, Forcefully Stop.
- To force the instance to stop using the AWS CLI, use the stop-instances command and the --force option as follows:

```
aws ec2 stop-instances --instance-ids i-0123ab456c789d01e --force
```

If, after 10 minutes, the instance has not stopped, post a request for help in the Amazon EC2 forum. To help expedite a resolution, include the instance ID, and describe the steps that you've already taken. Alternatively, if you have a support plan, create a technical support case in the Support Center.

Creating a Replacement Instance

To attempt to resolve the problem while you are waiting for assistance from the Amazon EC2 forum or the Support Center, create a replacement instance. Create an AMI of the stuck instance, and launch a new instance using the new AMI.

To create a replacement instance using the console

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances and select the stuck instance.
4. In the Create Image dialog box, fill in the following fields, and then choose Create Image:
   a. Specify a name and description for the AMI.
   b. Choose No reboot.

   For more information, see Creating a Windows AMI from a Running Instance (p. 65).
5. Launch a new instance from the AMI and verify that the new instance is working.
6. Select the stuck instance, and choose Actions, Instance State, Terminate. If the instance also gets stuck terminating, Amazon EC2 automatically forces it to terminate within a few hours.

To create a replacement instance using the CLI

1. Create an AMI from the stuck instance using the create-image (AWS CLI) command and the --no-reboot option as follows:
Terminating Your Instance

aws ec2 create-image --instance-id i-0123ab456c789d01e --name "AMI" --description "AMI for replacement instance" --no-reboot

2. Launch a new instance from the AMI using the run-instances (AWS CLI) command as follows:

aws ec2 run-instances --image-id ami-1a2b3c4d --count 1 --instance-type c3.large --key-name MyKeyPair --security-groups MySecurityGroup

3. Verify that the new instance is working.
4. Terminate the stuck instance using the terminate-instances (AWS CLI) command as follows:

aws ec2 terminate-instances --instance-ids i-1234567890abcdef0

If you are unable to create an AMI from the instance as described in the previous procedures, you can set up a replacement instance as follows:

(Alternate) To create a replacement instance using the console

1. Select the instance and choose Description, Block devices. Select each volume and write down its volume ID. Be sure to note which volume is the root volume.
2. In the navigation pane, choose Volumes. Select each volume for the instance, and choose Actions, Create Snapshot.
3. In the navigation pane, choose Snapshots. Select the snapshot that you just created, and choose Actions, Create Volume.
4. Launch an instance with the same operating system as the stuck instance. Note the volume ID and device name of its root volume.
5. In the navigation pane, choose Instances, select the instance that you just launched, choose Actions, Instance State, and then choose Stop.
6. In the navigation pane, choose Volumes, select the root volume of the stopped instance, and choose Actions, Detach Volume.
7. Select the root volume that you created from the stuck instance, choose Actions, Attach Volume, and attach it to the new instance as its root volume (using the device name that you wrote down). Attach any additional non-root volumes to the instance.
8. In the navigation pane, choose Instances and select the replacement instance. Choose Actions, Instance State, Start. Verify that the instance is working.
9. Select the stuck instance, choose Actions, Instance State, Terminate. If the instance also gets stuck terminating, Amazon EC2 automatically forces it to terminate within a few hours.

Troubleshooting Terminating (Shutting Down) Your Instance

You are not billed for any instance usage while an instance is not in the running state. In other words, when you terminate an instance, you stop incurring charges for that instance as soon as its state changes to shutting-down.

Delayed Instance Termination

If your instance remains in the shutting-down state longer than a few minutes, it might be delayed due to shutdown scripts being run by the instance.
Another possible cause is a problem with the underlying host computer. If your instance remains in the **shutting-down** state for several hours, Amazon EC2 treats it as a stuck instance and forcibly terminates it.

If it appears that your instance is stuck terminating and it has been longer than several hours, post a request for help to the Amazon EC2 forum. To help expedite a resolution, include the instance ID and describe the steps that you've already taken. Alternatively, if you have a support plan, create a technical support case in the Support Center.

**Terminated Instance Still Displayed**

After you terminate an instance, it remains visible for a short while before being deleted. The state shows as **terminated**. If the entry is not deleted after several hours, contact Support.

**Automatically Launch or Terminate Instances**

If you terminate all your instances, you may see that we launch a new instance for you. If you launch an instance, you may see that we terminate one of your instances. If you stop an instance, you may see that we terminate the instance and launch a new instance. Generally, these behaviors mean that you've used Amazon EC2 Auto Scaling or Elastic Beanstalk to scale your computing resources automatically based on criteria that you've defined.

For more information, see the Amazon EC2 Auto Scaling User Guide or the AWS Elastic Beanstalk Developer Guide.

**Sending a Diagnostic Interrupt (Advanced Users Only)**

**Warning**

Diagnostic interrupts are intended for use by advanced users. Incorrect usage could negatively impact your instance. Sending a diagnostic interrupt to an instance could trigger an instance to crash and reboot, which could lead to the loss of data.

You can send a diagnostic interrupt to an unreachable or unresponsive Windows instance to manually trigger a **stop error**. Stop errors are commonly referred to as **blue screen errors**.

In general, Windows operating systems crash and reboot when a stop error occurs, but the specific behavior depends on its configuration. A stop error can also cause the operating system to write debugging information, such as a kernel memory dump, to a file. You can then use this information to conduct root cause analysis to debug the instance.

The memory dump data is generated locally by the operating system on the instance itself.

Before sending a diagnostic interrupt to your instance, we recommend that you consult the documentation for your operating system and then make the necessary configuration changes.

**Contents**

- Supported Instance Types (p. 1069)
- Prerequisites (p. 1069)
- Sending a Diagnostic Interrupt (p. 1069)
Supported Instance Types

Diagnostic interrupt is supported on all Nitro-based instance types, except A1. For more information, see Nitro-based Instances (p. 122).

Prerequisites

Before using a diagnostic interrupt, you should configure your instance's operating system to perform the actions you need when a stop error occurs.

To configure Windows to generate a memory dump when a stop error occurs

1. Connect to your instance.
2. Open the Control Panel and choose System, Advanced system settings.
3. In the System Properties dialog box, choose the Advanced tab.
4. In the Startup and Recovery section, choose Settings...
5. In the System failure section, configure the settings as needed, and then choose OK.

For more information about configuring Windows stop errors, see Overview of memory dump file options for Windows.

Sending a Diagnostic Interrupt

After you have completed the necessary configuration changes, you can send a diagnostic interrupt to your instance using the AWS CLI or Amazon EC2 API.

To send a diagnostic interrupt to your instance (AWS CLI)

Use the send-diagnostic-interrupt command and specify the instance ID.

```
aws ec2 send-diagnostic-interrupt --instance-id i-1234567890abcdef0
```

To send a diagnostic interrupt to your instance (AWS Tools for Windows PowerShell)

Use the Send-EC2DiagnosticInterrupt cmdlt and specify the instance ID.

```
PS C:\> Send-EC2DiagnosticInterrupt-InstanceId i-1234567890abcdef0
```

Common Issues

The following are troubleshooting tips to help you solve common issues with EC2 instance running Windows Server.

Issues

- EBS Volumes Don’t Initialize on Windows Server 2016 and Later AMIs (p. 1070)
- Boot an EC2 Windows Instance into Directory Services Restore Mode (DSRM) (p. 1070)
- Instance loses network connectivity or scheduled tasks don’t run when expected (p. 1072)
- Unable to get console output (p. 1072)
- Windows Server 2012 R2 not available on the network (p. 1073)
EBS Volumes Don't Initialize on Windows Server 2016 and Later AMIs

Instances created from Windows Server 2016 and later Amazon Machine Images (AMIs) use the EC2Launch service for a variety of startup tasks, including initializing EBS volumes. By default, EC2Launch does not initialize secondary volumes. You can configure EC2Launch to initialize these disks automatically.

To map drive letters to volumes

1. Connect to the instance to configure and open the C:\ProgramData\Amazon\EC2-Windows\Launch\Config\DriveLetterMappingConfig.json file in a text editor.
2. Specify the volume settings using the following format:

   ```json
   {
   "driveLetterMapping": [
   {
     "volumeName": "sample volume",
     "driveLetter": "H"
   }
   ]
   }
   ```

3. Save your changes and close the file.
4. Open Windows PowerShell and use the following command to run the EC2Launch script that initializes the disks:

   ```powershell
   PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeDisks.ps1
   ```

To initialize the disks each time the instance boots, add the -Schedule flag as follows:

   ```powershell
   PS C:\> C:\ProgramData\Amazon\EC2-Windows\Launch\Scripts\InitializeDisks.ps1 -Schedule
   ```

Boot an EC2 Windows Instance into Directory Services Restore Mode (DSRM)

If an instance running Microsoft Active Directory experiences a system failure or other critical issues you can troubleshoot the instance by booting into a special version of Safe Mode called Directory Services Restore Mode (DSRM). In DSRM you can repair or recover Active Directory.

Driver Support for DSRM

How you enable DSRM and boot into the instance depends on the drivers the instance is running. In the EC2 console you can view driver version details for an instance from the System Log. The following table shows which drivers are supported for DSRM.

<table>
<thead>
<tr>
<th>Driver Versions</th>
<th>DSRM Supported?</th>
<th>Next Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrix PV 5.9</td>
<td>No</td>
<td>Restore the instance from a backup. You cannot enable DSRM.</td>
</tr>
</tbody>
</table>
Driver Versions | DSRM Supported? | Next Steps
---|---|---
AWS PV 7.2.0 | No | Though DSRM is not supported for this driver, you can still detach the root volume from the instance, take a snapshot of the volume or create an AMI from it, and attach it to another instance in the same Availability Zone as a secondary volume. You can then enable DSRM (as described in this section).
AWS PV 7.2.2 and later | Yes | Detach the root volume, attach it to another instance, and enable DSRM (as described in this section).
Enhanced Networking | Yes | Detach the root volume, attach it to another instance, and enable DSRM (as described in this section).

For information about how to enable Enhanced Networking, see Enabling Enhanced Networking on Windows Instances in a VPC. For more information about upgrading AWS PV drivers, see Upgrading PV Drivers on Your Windows Instances (p. 436).

Configure an Instance to Boot into DSRM

EC2 Windows instances do not have network connectivity before the operating system is running. For this reason, you cannot press the F8 button on your keyboard to select a boot option. You must use one of the following procedures to boot an EC2 Windows Server instance into DSRM.

If you suspect that Active Directory has been corrupted and the instance is still running, you can configure the instance to boot into DSRM using either the System Configuration dialog box or the command prompt.

To boot an online instance into DSRM using the System Configuration dialog box

1. In the Run dialog box, type `msconfig` and press Enter.
2. Choose the Boot tab.
4. Choose Active Directory repair and then choose OK. The system prompts you to reboot the server.

To boot an online instance into DSRM using the command line

From a Command Prompt window, run the following command:

```
bcdedit /set safeboot dsrepair
```

If an instance is offline and unreachable, you must detach the root volume and attach it to another instance to enable DSRM mode.

To boot an offline instance into DSRM

1. Open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the navigation pane, choose Instances.
3. Locate the affected instance. Open the context (right-click) menu for the instance, choose Instance State, and then choose Stop.
4. Choose Launch Instance and create a temporary instance in the same Availability Zone as the affected instance. Choose an instance type that uses a different version of Windows. For example, if your instance is Windows Server 2008, then choose a Windows Server 2008 R2 instance.
Important
If you do not create the instance in the same Availability Zone as the affected instance you will not be able to attach the root volume of the affected instance to the new instance.

5. In the navigation pane, choose Volumes.
6. Locate the root volume of the affected instance. Detach the volume and attach it to the temporary instance you created earlier. Attach it with the default device name (xvdf).
7. Use Remote Desktop to connect to the temporary instance, and then use the Disk Management utility to make the volume available for use.
8. Open a command prompt and run the following command. Replace D with the actual drive letter of the secondary volume you just attached:

   \bcedit /store D:\Boot\BCD /set (default) safeboot dsrepair

9. In the Disk Management Utility, choose the drive you attached earlier, open the context (right-click) menu, and choose Offline.
10. In the EC2 console, detach the affected volume from the temporary instance and reattach it to your original instance with the device name /dev/sda1. You must specify this device name to designate the volume as a root volume.
11. Start the instance.
12. After the instance passes the health checks in the EC2 console, connect to the instance using Remote Desktop and verify that it boots into DSRM mode.
13. (Optional) Delete or stop the temporary instance you created in this procedure.

Instance loses network connectivity or scheduled tasks don't run when expected

If you restart your instance and it loses network connectivity, it's possible that the instance has the wrong time.

By default, Windows instances use Coordinated Universal Time (UTC). If you set the time for your instance to a different time zone and then restart it, the time becomes offset and the instance temporarily loses its IP address. The instance regains network connectivity eventually, but this can take several hours. The amount of time that it takes for the instance to regain network connectivity depends on the difference between UTC and the other time zone.

This same time issue can also result in scheduled tasks not running when you expect them to. In this case, the scheduled tasks do not run when expected because the instance has the incorrect time.

To use a time zone other than UTC persistently, you must set the \RealTimeIsUniversal\ registry key. Without this key, an instance uses UTC after you restart it.

To resolve time issues that cause a loss of network connectivity

1. Ensure that you are running the recommended PV drivers. For more information, see Upgrading PV Drivers on Your Windows Instances (p. 436).
2. Verify that the following registry key exists and is set to 1: HKEY_LOCAL_MACHINE\SYSTEM \CurrentControlSet\Control\TimeZoneInformation\RealTimeIsUniversal

Unable to get console output

For Windows instances, the instance console displays the output from tasks performed during the Windows boot process. If Windows boots successfully, the last message logged is Windows is Ready

1072
Windows Server 2012 R2 not available on the network

For information about troubleshooting a Windows Server 2012 R2 instance that is not available on the network, see Windows Server 2012 R2 loses network and storage connectivity after an instance reboot (p. 441).

Common Messages

This section includes tips to help you troubleshoot issues based on common messages.

Topics
- "Password is not available" (p. 1073)
- "Password not available yet" (p. 1074)
- "Cannot retrieve Windows password" (p. 1074)
- "Waiting for the metadata service" (p. 1074)
- "Unable to activate Windows" (p. 1077)
- "Windows is not genuine (0x80070005)" (p. 1078)
- "No Terminal Server License Servers available to provide a license" (p. 1078)
- "Some settings are managed by your organization" (Windows 2019) (p. 1079)

"Password is not available"

To connect to a Windows instance using Remote Desktop, you must specify an account and password. The accounts and passwords provided are based on the AMI that you used to launch the instance. You can either retrieve the auto-generated password for the Administrator account, or use the account and password that were in use in the original instance from which the AMI was created.

If your Windows instance isn't configured to generate a random password, you'll receive the following message when you retrieve the auto-generated password using the console:

Password is not available.
The instance was launched from a custom AMI, or the default password has changed. A password cannot be retrieved for this instance. If you have forgotten your password, you can reset it using the Amazon EC2 configuration service. For more information, see Passwords for a Windows Server instance.
Check the console output for the instance to see whether the AMI that you used to launch it was created with password generation disabled. If password generation is disabled, the console output contains the following:

```
EC2SetPassword: Disabled
```

If password generation is disabled and you don't remember the password for the original instance, you can reset the password for this instance. For more information, see Resetting a Lost or Expired Windows Administrator Password (p. 1058).

"Password not available yet"

To connect to a Windows instance using Remote Desktop, you must specify an account and password. The accounts and passwords provided are based on the AMI that you used to launch the instance. You can either retrieve the auto-generated password for the Administrator account, or use the account and password that were in use in the original instance from which the AMI was created.

Your password should be available within a few minutes. If the password isn't available, you'll receive the following message when you retrieve the auto-generated password using the console:

```
Password not available yet.
Please wait at least 4 minutes after launching an instance before trying to retrieve the auto-generated password.
```

If it's been longer than four minutes and you still can't get the password, it's possible that EC2Config is disabled. Verify by checking whether the console output is empty. For more information, see Unable to get console output (p. 1072).

Also verify that the AWS Identity and Access Management (IAM) account being used to access the Management Portal has the ec2:GetPasswordData action allowed. For more information about IAM permissions, see What is IAM?.

"Cannot retrieve Windows password"

To retrieve the auto-generated password for the Administrator account, you must use the private key for the key pair that you specified when you launched the instance. If you didn't specify a key pair when you launched the instance, you'll receive the following message.

```
Cannot retrieve Windows password
```

You can terminate this instance and launch a new instance using the same AMI, making sure to specify a key pair.

"Waiting for the metadata service"

A Windows instance must obtain information from its instance metadata before it can activate itself. By default, the WaitForMetaDataAvailable setting ensures that the EC2Config service waits for the instance metadata to be accessible before continuing with the boot process. For more information, see Instance Metadata and User Data (p. 478).

If the instance is failing the instance reachability test, try the following to resolve this issue.

- Check the CIDR block for your VPC. A Windows instance cannot boot correctly if it's launched into a VPC that has an IP address range from 224.0.0.0 to 255.255.255.255 (Class D and Class E IP address ranges). These IP address ranges are reserved, and should not be assigned to host devices.
We recommend that you create a VPC with a CIDR block from the private (non-publicly routable) IP address ranges as specified in RFC 1918.

- It's possible that the system has been configured with a static IP address. Try creating a network interface (p. 691) and attaching it to the instance (p. 693).

To enable DHCP on a Windows instance that you can't connect to

1. Stop the affected instance and detach its root volume.
2. Launch a temporary instance in the same Availability Zone as the affected instance.
   
   **Warning**
   
   If your temporary instance is based on the same AMI that the original instance is based on, you must complete additional steps or you won't be able to boot the original instance after you restore its root volume because of a disk signature collision. Alternatively, select a different AMI for the temporary instance. For example, if the original instance uses the AWS Windows AMI for Windows Server 2008 R2, launch the temporary instance using the AWS Windows AMI for Windows Server 2012.
   
   3. Attach the root volume from the affected instance to this temporary instance. Connect to the temporary instance, open the **Disk Management** utility, and bring the drive online.
   
   4. From the temporary instance, open **Regedit** and select **HKEY_LOCAL_MACHINE**. From the **File** menu, choose **Load Hive**. Select the drive, open the file **Windows\System32\config\SYSTEM**, and specify a key name when prompted (you can use any name).
   
   5. Select the key that you just loaded and navigate to **ControlSet001\Services\Tcpip\Parameters\Interfaces**. Each network interface is listed by a GUID. Select the correct network interface. If DHCP is disabled and a static IP address assigned, **EnableDHCP** is set to 0. To enable DHCP, set **EnableDHCP** to 1, and delete the following keys if they exist: **NameServer**, **SubnetMask**, **IPAddress**, and **DefaultGateway**. Select the key again, and from the **File** menu, choose **Unload Hive**.

   **Note**
   
   If you have multiple network interfaces, you'll need to identify the correct interface to enable DHCP. To identify the correct network interface, review the following key values **NameServer**, **SubnetMask**, **IPAddress**, and **DefaultGateway**. These values display the static configuration of the previous instance.

6. (Optional) If DHCP is already enabled, it's possible that you don't have a route to the metadata service. Updating EC2Config can resolve this issue.
   
   a. **Download** and install the latest version of the EC2Config service. For more information about installing this service, see **Installing the Latest Version of EC2Config** (p. 406).
   
   b. Extract the files from the .zip file to the Temp directory on the drive you attached.
   
   c. Open **Regedit** and select **HKEY_LOCAL_MACHINE**. From the **File** menu, choose **Load Hive**. Select the drive, open the file **Windows\System32\config\SOFTWARE**, and specify a key name when prompted (you can use any name).
   
   d. Select the key that you just loaded and navigate to **Microsoft\Windows\CurrentVersion**. Select the RunOnce key. (If this key doesn't exist, right-click **CurrentVersion**, point to **New**, select **Key**, and name the key **RunOnce**.) Right-click, point to **New**, and select **String Value**. Enter **Ec2Install** as the name and **C:\Temp\Ec2Install.exe -q** as the data.
   
   e. Select the key again, and from the **File** menu, choose **Unload Hive**.

7. (Optional) If your temporary instance is based on the same AMI that the original instance is based on, you must complete the following steps or you won't be able to boot the original instance after you restore its root volume because of a disk signature collision.

   **Warning**
   
   The following procedure describes how to edit the Windows Registry using Registry Editor. If you are not familiar with the Windows Registry or how to safely make changes using Registry Editor, see **Configure the Registry**.
a. Open a command prompt, type `regedit.exe`, and press Enter.

b. In the Registry Editor, choose HKEY_LOCAL_MACHINE from the context menu (right-click), and then choose Find.

c. Type Windows Boot Manager and then choose Find Next.

d. Choose the key named 11000001. This key is a sibling of the key you found in the previous step.

e. In the right pane, choose Element and then choose Modify from the context menu (right-click).

f. Locate the four-byte disk signature at offset 0x38 in the data. Reverse the bytes to create the disk signature, and write it down. For example, the disk signature represented by the following data is E9EB3AA5:

```
... 0030 00 00 00 00 01 00 00 00 00 38 A5 3A EB E9 00 00 00 00 00 00 00 00 00 00 00 ...
```

g. In a Command Prompt window, run the following command to start Microsoft DiskPart.

```
 diskpart
```

h. Run the following DiskPart command to select the volume. (You can verify that the disk number is 1 using the Disk Management utility.)

```
DISKPART> select disk 1
Disk 1 is now the selected disk.
```

i. Run the following DiskPart command to get the disk signature.

```
DISKPART> uniqueid disk
Disk ID: OC764FA8
```

j. If the disk signature shown in the previous step doesn't match the disk signature from BCD that you wrote down earlier, use the following DiskPart command to change the disk signature so that it matches:

```
DISKPART> uniqueid disk id=E9EB3AA5
```

8. Using the Disk Management utility, bring the drive offline.

   **Note**
   
The drive is automatically offline if the temporary instance is running the same operating system as the affected instance, so you won't need to bring it offline manually.

9. Detach the volume from the temporary instance. You can terminate the temporary instance if you have no further use for it.

10. Restore the root volume of the affected instance by attaching the volume as `/dev/sda1`.

11. Start the affected instance.

If you are connected to the instance, open an Internet browser from the instance and enter the following URL for the metadata server:
If you can't contact the metadata server, try the following to resolve the issue:

- **Download** and install the latest version of the EC2Config service. For more information about installing this service, see *Installing the Latest Version of EC2Config* (p. 406).
- **Check** whether the Windows instance is running RedHat PV drivers. If so, update to Citrix PV drivers. For more information, see *Upgrading PV Drivers on Your Windows Instances* (p. 436).
- **Verify** that the firewall, IPSec, and proxy settings do not block outgoing traffic to the metadata service (169.254.169.254) or the KMS servers (the addresses are specified in TargetKMSServer elements in C:\Program Files\Amazon\Ec2ConfigService\Settings\ActivationSettings.xml).
- **Verify** that you have a route to the metadata service (169.254.169.254) using the following command.

```
route print
```

- **Check** for network issues that might affect the Availability Zone for your instance. Go to [http://status.aws.amazon.com/](http://status.aws.amazon.com/).

"Unable to activate Windows"

Windows instances use Windows KMS activation. You can receive this message: A problem occurred when Windows tried to activate. Error Code 0xC004F074, if your instance can't reach the KMS server. Windows must be activated every 180 days. EC2Config attempts to contact the KMS server before the activation period expires to ensure that Windows remains activated.

If you encounter a Windows activation issue, use the following procedure to resolve the issue.

**For EC2Config (Windows Server 2012 R2 AMIs and earlier)**

1. **Download** and install the latest version of the EC2Config service. For more information about installing this service, see *Installing the Latest Version of EC2Config* (p. 406).
2. Log onto the instance and open the following file: C:\Program Files\Amazon\Ec2ConfigService\Settings\config.xml.
3. Locate the **Ec2WindowsActivate** plugin in the **config.xml** file. Change the state to **Enabled** and save your changes.
4. In the Windows Services snap-in, restart the EC2Config service or reboot the instance.

If this does not resolve the activation issue, follow these additional steps.

1. Set the KMS target: `C:\> slmgr.vbs /skms 169.254.169.250:1688`
2. Activate Windows: `C:\> slmgr.vbs /ato`

**For EC2Launch (Windows Server 2016 AMIs and later)**

1. **Import** the EC2Launch module:

```
PS C:\> Import-Module "C:\ProgramData\Amazon\EC2-Windows\Launch\Module\Ec2Launch.psd1"
```
2. **Call** the Add-Routes function:

```
PS C:\> Add-Routes
```
3. Call the Set-ActivationSettings function:

```
PS C:\> Set-ActivationSettings
```

4. Then, run the following script to activate Windows:

```
PS C:\> cscript "#{env:SYSTEMROOT}\system32\slmgr.vbs" /ato
```

For both EC2Config and EC2Launch, if you are still receiving an activation error, verify the following information.

- Verify that you have routes to the KMS servers. Open C:\Program Files\Amazon\Ec2ConfigService\Settings\ActivationSettings.xml and locate the TargetKMServer elements. Run the following command and check whether the addresses for these KMS servers are listed.

```
route print
```

- Verify that the KMS client key is set. Run the following command and check the output.

```
C:\Windows\System32\slmgr.vbs /dlv
```

If the output contains Error: product key not found, the KMS client key isn't set. If the KMS client key isn't set, look up the client key as described in this Microsoft article: KMS Client Setup Keys, and then run the following command to set the KMS client key.

```
C:\Windows\System32\slmgr.vbs /ipk client_key
```

- Verify that the system has the correct time and time zone. If you are using Windows Server 2008 or later and a time zone other than UTC, add the following registry key and set it to 1 to ensure that the time is correct:

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Control\TimeZoneInformation\RealTimeIsUniversal.
```

- If Windows Firewall is enabled, temporarily disable it using the following command.

```
netsh advfirewall set allprofiles state off
```

"Windows is not genuine (0x80070005)"

Windows instances use Windows KMS activation. If an instance is unable to complete the activation process, it reports that the copy of Windows is not genuine.

Try the suggestions for "Unable to activate Windows" (p. 1077).

"No Terminal Server License Servers available to provide a license"

By default, Windows Server is licensed for two simultaneous users through Remote Desktop. If you need to provide more than two users with simultaneous access to your Windows instance through Remote Desktop, you can purchase a Remote Desktop Services client access license (CAL) and install the Remote Desktop Session Host and Remote Desktop Licensing Server roles.

Check for the following issues:
• You've exceeded the maximum number of concurrent RDP sessions.
• You've installed the Windows Remote Desktop Services role.
• Licensing has expired. If the licensing has expired, you can't connect to your Windows instance as a user. You can try the following:
  • Connect to the instance from the command line using an `/admin` parameter, for example:

```
mstsc /v:instance /admin
```

For more information, see the following Microsoft article: Access Remote Desktop Via Command Line.
• Stop the instance, detach its Amazon EBS volumes, and attach them to another instance in the same Availability Zone to recover your data.

"Some settings are managed by your organization" (Windows 2019)

Instances launched from the latest Windows Server 2019 AMIs may show a Windows Update dialog message stating "Some settings are managed by your organization." This message appears as a result of changes in Windows Server 2019 and does not impact the behavior of Windows Update or your ability to manage update settings.

To remove this warning, follow these steps:

1. In gpedit.msc, navigate to Computer Configuration > Administrative Templates > Windows Components > Windows updates. Edit Configure Automatic Update, and switch this to enabled.
2. In a command prompt, update group policy with `gpupdate /force`.
3. Close and reopen the Windows Update Settings. You will see the above message about your settings being managed by your organization, followed by "We'll automatically download updates, except on metered connections (where charges may apply). In that case, we'll automatically download those updates required to keep Windows running smoothly.
4. Return to gpedit.msc in Step 1 and set the group policy back to not configured. Do another `gpupdate /force`.
5. Close the command prompt and wait a few minutes.
6. Reopen the Windows Update Settings. No message stating "Some settings are managed by your organization" should appear.
# Document History

The following table describes important additions to the Amazon EC2 documentation. We also update the documentation frequently to address the feedback that you send us.

**Current API version: 2016-11-15**

<table>
<thead>
<tr>
<th>Feature</th>
<th>API Version</th>
<th>Description</th>
<th>Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>G4 instances</td>
<td>2016-11-15</td>
<td>G4 instances use NVIDIA Tesla GPUs and provide a cost-effective, high-performance platform for general purpose GPU computing using the CUDA or machine learning frameworks along with graphics applications using DirectX or OpenGL.</td>
<td>19 September 2019</td>
</tr>
<tr>
<td>Diagnostic interrupt</td>
<td>2016-11-15</td>
<td>You can send a diagnostic interrupt to an unreachable or unresponsive instance to trigger a kernel panic (on Linux instances), or a blue screen/stop error (on Windows instances). For more information, see Sending a Diagnostic Interrupt (Advanced Users Only) (p. 1068).</td>
<td>14 August 2019</td>
</tr>
<tr>
<td>Capacity optimized allocation strategy</td>
<td>2016-11-15</td>
<td>Using EC2 Fleet or Spot Fleet, you can now launch Spot Instances from Spot pools with optimal capacity for the number of instances that are launching. For more information, see Configuring EC2 Fleet for Capacity Optimization (p. 355).</td>
<td>12 August 2019</td>
</tr>
<tr>
<td>On-Demand Capacity Reservation</td>
<td>2016-11-15</td>
<td>You can now share your Capacity Reservations across AWS accounts. For more information, see Working with Shared Capacity Reservations (p. 324).</td>
<td>29 July 2019</td>
</tr>
<tr>
<td>Tag launch templates on creation</td>
<td>2016-11-15</td>
<td>You can tag launch templates at creation. For more information, see Tagging Your Resources (p. 966).</td>
<td>24 July 2019</td>
</tr>
<tr>
<td>Maximum total price</td>
<td>2016-11-15</td>
<td>You can specify a maximum hourly price for all On-Demand Instances and Spot Instances in both EC2 Fleet and Spot Fleet. For more information, see Control Spending (p. 357) in EC2 Fleet and Control Spending (p. 238) in Spot Fleet.</td>
<td>1 July 2019</td>
</tr>
<tr>
<td>Host recovery</td>
<td>2016-11-15</td>
<td>Automatically restart your instances on a new host in the event of an unexpected hardware failure on a Dedicated Host. For more information, see Host Recovery (p. 307).</td>
<td>5 June 2019</td>
</tr>
<tr>
<td>Amazon EBS multi-volume snapshots</td>
<td>2016-11-15</td>
<td>Take exact point-in-time, data coordinated, and crash-consistent snapshots across multiple EBS volumes attached to an EC2 instance.</td>
<td>29 May 2019</td>
</tr>
<tr>
<td>Amazon EBS encryption by default</td>
<td>2016-11-15</td>
<td>After you enable encryption by default in a Region, all new EBS volumes you create in the Region are encrypted using the default CMK.</td>
<td>23 May 2019</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>VSS Application-Consistent Snapshots</td>
<td>2016-11-15</td>
<td>Take application-consistent snapshots of all Amazon EBS volumes attached to your Amazon EC2 Windows instances by using Run Command. For more information, see Encryption by Default (p. 868).</td>
<td>13 May 2019</td>
</tr>
<tr>
<td>Tag VPC endpoints, endpoint services, and endpoint service configurations</td>
<td>2016-11-15</td>
<td>You can tag VPC endpoints, endpoint services, and endpoint service configurations. For more information, see Tagging Your Resources (p. 966).</td>
<td>13 May 2019</td>
</tr>
<tr>
<td>Windows to Linux Replatforming Assistant for Microsoft SQL Server Databases</td>
<td>2016-11-15</td>
<td>Move existing Microsoft SQL Server workloads from a Windows to a Linux operating system.</td>
<td>8 May 2019</td>
</tr>
<tr>
<td>I3en instances</td>
<td>2016-11-15</td>
<td>New instances featuring up to 100 Gbps of network bandwidth.</td>
<td>8 May 2019</td>
</tr>
<tr>
<td>Windows Automated Upgrade</td>
<td>2016-11-15</td>
<td>Perform automated upgrade on Windows Server 2008 R2 and SQL Server 2008 R2 EC2 instances using AWS Systems Manager SSM documents.</td>
<td>6 May 2019</td>
</tr>
<tr>
<td>T3a instances</td>
<td>2016-11-15</td>
<td>New instances featuring AMD EYPC processors.</td>
<td>24 April 2019</td>
</tr>
<tr>
<td>M5ad and R5ad instances</td>
<td>2016-11-15</td>
<td>New instances featuring AMD EYPC processors.</td>
<td>27 March 2019</td>
</tr>
<tr>
<td>Tag Dedicated Host Reservations</td>
<td>2016-11-15</td>
<td>You can tag your Dedicated Host Reservations. For more information, see Tagging Dedicated Host Reservations (p. 307).</td>
<td>14 March 2019</td>
</tr>
<tr>
<td>Bare metal instances for M5, M5d, R5, R5d, and z1d</td>
<td>2016-11-15</td>
<td>New instances that provide your applications with direct access to the physical resources of the host server.</td>
<td>13 February 2019</td>
</tr>
<tr>
<td>Partition placement groups</td>
<td>2016-11-15</td>
<td>Partition placement groups spread instances across logical partitions, ensuring that instances in one partition do not share underlying hardware with instances in other partitions. For more information, see Partition Placement Groups (p. 711).</td>
<td>20 December 2018</td>
</tr>
<tr>
<td>p3dn.24xlarge instances</td>
<td>2016-11-15</td>
<td>New p3dn.24xlarge instances provide 100 Gbps of network bandwidth.</td>
<td>7 December 2018</td>
</tr>
<tr>
<td>Instances featuring 100 Gbps of network bandwidth</td>
<td>2016-11-15</td>
<td>New C5n instances can utilize up to 100 Gbps of network bandwidth.</td>
<td>26 November 2018</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Spot console recommends a fleet of instances</td>
<td>2016-11-15</td>
<td>The Spot console recommends a fleet of instances based on Spot best practice (instance diversification) to meet the minimum hardware specifications (vCPUs, memory, and storage) for your application need. For more information, see Creating a Spot Fleet Request (p. 259).</td>
<td>20 November 2018</td>
</tr>
<tr>
<td>New EC2 Fleet request type: instant</td>
<td>2016-11-15</td>
<td>EC2 Fleet now supports a new request type, instant, that you can use to synchronously provision capacity across instance types and purchase models. The instant request returns the launched instances in the API response, and takes no further action, enabling you to control if and when instances are launched. For more information, see EC2 Fleet Request Types (p. 354).</td>
<td>14 November 2018</td>
</tr>
<tr>
<td>Instances featuring AMD EYPC processors</td>
<td>2016-11-15</td>
<td>New general purpose (M5a) and memory optimized instances (R5a) offer lower-priced options for microservices, small to medium databases, virtual desktops, development and test environments, business applications, and more.</td>
<td>6 November 2018</td>
</tr>
<tr>
<td>Spot savings information</td>
<td>2016-11-15</td>
<td>You can view the savings made from using Spot Instances for a single Spot Fleet or for all Spot Instances. For more information, see Savings From Purchasing Spot Instances (p. 243).</td>
<td>5 November 2018</td>
</tr>
<tr>
<td>Console support for optimizing CPU options</td>
<td>2016-11-15</td>
<td>When you launch an instance, you can optimize the CPU options to suit specific workloads or business needs using the Amazon EC2 console. For more information, see Optimizing CPU Options (p. 447).</td>
<td>31 October 2018</td>
</tr>
<tr>
<td>Console support for creating a launch template from an instance</td>
<td>2016-11-15</td>
<td>You can create a launch template using an instance as the basis for a new launch template using the Amazon EC2 console. For more information, see Creating a Launch Template (p. 340).</td>
<td>30 October 2018</td>
</tr>
<tr>
<td>On-Demand Capacity Reservations</td>
<td>2016-11-15</td>
<td>You can reserve capacity for your Amazon EC2 instances in a specific Availability Zone for any duration. This allows you to create and manage capacity reservations independently from the billing discounts offered by Reserved Instances (RI). For more information, see On-Demand Capacity Reservations (p. 317).</td>
<td>25 October 2018</td>
</tr>
<tr>
<td>Bring Your Own IP Addresses (BYOIP)</td>
<td>2016-11-15</td>
<td>You can bring part or all of your public IPv4 address range from your on-premises network to your AWS account. After you bring the address range to AWS, it appears in your account as an address pool. You can create an Elastic IP address from your address pool and use it with your AWS resources. For more information, see Bring Your Own IP Addresses (BYOIP) (p. 670).</td>
<td>23 October 2018</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>g3s.xlarge instances</td>
<td>2016-11-15</td>
<td>Expands the range of the accelerated-computing G3 instance family with the introduction of g3s.xlarge instances.</td>
<td>11 October 2018</td>
</tr>
<tr>
<td>Dedicated Host tag on create and console support</td>
<td>2016-11-15</td>
<td>You can tag your Dedicated Hosts on creation, and you can manage your Dedicated Host tags using the Amazon EC2 console. For more information, see Allocating Dedicated Hosts (p. 298).</td>
<td>08 October 2018</td>
</tr>
<tr>
<td>High memory instances</td>
<td>2016-11-15</td>
<td>These instances are purpose-built to run large in-memory databases. They offer bare metal performance with direct access to host hardware. For more information, see Memory Optimized Instances (p. 165).</td>
<td>27 September 2018</td>
</tr>
<tr>
<td>f1.4xlarge instances</td>
<td>2016-11-15</td>
<td>Expands the range of the accelerated-computing F1 instance family with the introduction of f1.4xlarge instances.</td>
<td>25 September 2018</td>
</tr>
<tr>
<td>Console support for scheduled scaling for Spot Fleet</td>
<td>2016-11-15</td>
<td>Increase or decrease the current capacity of the fleet based on the date and time. For more information, see Scale Spot Fleet Using Scheduled Scaling (p. 278).</td>
<td>20 September 2018</td>
</tr>
<tr>
<td>T3 instances</td>
<td>2016-11-15</td>
<td>T3 instances are the next generation burstable general-purpose instance type that provide a baseline level of CPU performance with the ability to burst CPU usage at any time for as long as required. For more information, see Burstable Performance Instances (p. 132).</td>
<td>21 August 2018</td>
</tr>
<tr>
<td>Allocation strategies for EC2 Fleets</td>
<td>2016-11-15</td>
<td>You can specify whether On-Demand capacity is fulfilled by price (lowest price first) or priority (highest priority first). You can specify the number of Spot pools across which to allocate your target Spot capacity. For more information, see Allocation Strategies for Spot Instances (p. 355).</td>
<td>26 July 2018</td>
</tr>
<tr>
<td>Allocation strategies for Spot Fleets</td>
<td>2016-11-15</td>
<td>You can specify whether On-Demand capacity is fulfilled by price (lowest price first) or priority (highest priority first). You can specify the number of Spot pools across which to allocate your target Spot capacity. For more information, see Allocation Strategy for Spot Instances (p. 236).</td>
<td>26 July 2018</td>
</tr>
<tr>
<td>R5 and R5d instances</td>
<td>2016-11-15</td>
<td>R5 and R5d instances are ideally suited for high-performance databases, distributed in-memory caches, and in-memory analytics. R5d instances come with NVMe instance store volumes. For more information, see Memory Optimized Instances (p. 165).</td>
<td>25 July 2018</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>z1d instances</td>
<td>2016-11-15</td>
<td>These instances are designed for applications that require high per-core performance with a large amount of memory, such as electronic design automation (EDA) and relational databases. These instances come with NVME instance store volumes. For more information, see Memory Optimized Instances (p. 165).</td>
<td>25 July 2018</td>
</tr>
<tr>
<td>Automate snapshot lifecycle</td>
<td>2016-11-15</td>
<td>You can use Amazon Data Lifecycle Manager to automate creation and deletion of snapshots for your EBS volumes. For more information, see Automating the Amazon EBS Snapshot Lifecycle (p. 844).</td>
<td>12 July 2018</td>
</tr>
<tr>
<td>Launch template CPU options</td>
<td>2016-11-15</td>
<td>When you create a launch template using the command line tools, you can optimize the CPU options to suit specific workloads or business needs. For more information, see Creating a Launch Template (p. 340).</td>
<td>11 July 2018</td>
</tr>
<tr>
<td>Tag Dedicated Hosts</td>
<td>2016-11-15</td>
<td>You can tag your Dedicated Hosts. For more information, see Tagging Dedicated Hosts (p. 302).</td>
<td>3 July 2018</td>
</tr>
<tr>
<td>i3.metal instances</td>
<td>2016-11-15</td>
<td>i3.metal instances provide your applications with direct access to the physical resources of the host server, such as processors and memory. For more information, see Storage Optimized Instances (p. 173).</td>
<td>17 May 2018</td>
</tr>
<tr>
<td>Get latest console output</td>
<td>2016-11-15</td>
<td>You can retrieve the latest console output for some instance types when you use the get-console-output AWS CLI command.</td>
<td>9 May 2018</td>
</tr>
<tr>
<td>Optimize CPU options</td>
<td>2016-11-15</td>
<td>When you launch an instance, you can optimize the CPU options to suit specific workloads or business needs. For more information, see Optimizing CPU Options (p. 447).</td>
<td>8 May 2018</td>
</tr>
<tr>
<td>EC2 Fleet</td>
<td>2016-11-15</td>
<td>You can use EC2 Fleet to launch a group of instances across different EC2 instance types and Availability Zones, and across On-Demand Instance, Reserved Instance, and Spot Instance purchasing models. For more information, see Launching an EC2 Fleet (p. 351).</td>
<td>2 May 2018</td>
</tr>
<tr>
<td>On-Demand Instances in Spot Fleets</td>
<td>2016-11-15</td>
<td>You can include a request for On-Demand capacity in your Spot Fleet request to ensure that you always have instance capacity. For more information, see How Spot Fleet Works (p. 235).</td>
<td>2 May 2018</td>
</tr>
<tr>
<td>Tag EBS snapshots on creation</td>
<td>2016-11-15</td>
<td>You can apply tags to snapshots during creation. For more information, see Creating Amazon EBS Snapshots (p. 822).</td>
<td>2 April 2018</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Change placement groups</td>
<td>2016-11-15</td>
<td>You can move an instance in or out of a placement group, or change its placement group. For more information, see Changing the Placement Group for an Instance (p. 717).</td>
<td>1 March 2018</td>
</tr>
<tr>
<td>Longer resource IDs</td>
<td>2016-11-15</td>
<td>You can enable the longer ID format for more resource types. For more information, see Resource IDs (p. 956).</td>
<td>9 February 2018</td>
</tr>
<tr>
<td>Network performance</td>
<td>2016-11-15</td>
<td>Instances outside of a cluster placement group can now benefit from increased bandwidth when sending or receiving network traffic between other instances or Amazon S3. For more information, see Networking and Storage Features (p. 123).</td>
<td>24 January 2018</td>
</tr>
<tr>
<td>improvements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tag Elastic IP addresses</td>
<td>2016-11-15</td>
<td>You can tag your Elastic IP addresses. For more information, see Tagging an Elastic IP Address (p. 676).</td>
<td>21 December 2017</td>
</tr>
<tr>
<td>Amazon Time Sync Service</td>
<td>2016-11-15</td>
<td>You can use the Amazon Time Sync Service to keep accurate time on your instance. For more information, see Setting the Time for a Windows Instance (p. 460).</td>
<td>29 November 2017</td>
</tr>
<tr>
<td>T2 Unlimited</td>
<td>2016-11-15</td>
<td>T2 Unlimited instances can burst above the baseline for as long as required. For more information, see Burstable Performance Instances (p. 132).</td>
<td>29 November 2017</td>
</tr>
<tr>
<td>Launch templates</td>
<td>2016-11-15</td>
<td>A launch template can contain all or some of the parameters to launch an instance, so that you don't have to specify them every time you launch an instance. For more information, see Launching an Instance from a Launch Template (p. 338).</td>
<td>29 November 2017</td>
</tr>
<tr>
<td>Spread placement</td>
<td>2016-11-15</td>
<td>Spread placement groups are recommended for applications that have a small number of critical instances that should be kept separate from each other. For more information, see Spread Placement Groups (p. 712).</td>
<td>29 November 2017</td>
</tr>
<tr>
<td>H1 instances</td>
<td>2016-11-15</td>
<td>H1 instances are designed for high-performance big data workloads. For more information, see Storage Optimized Instances (p. 173).</td>
<td>28 November 2017</td>
</tr>
<tr>
<td>M5 instances</td>
<td>2016-11-15</td>
<td>M5 instances are the next generation of general purpose compute instances. They provide a balance of compute, memory, storage, and network resources.</td>
<td>28 November 2017</td>
</tr>
<tr>
<td>Spot Instance hibernation</td>
<td>2016-11-15</td>
<td>The Spot service can hibernate Spot Instances in the event of an interruption. For more information, see Hibernating Interrupted Spot Instances (p. 287).</td>
<td>28 November 2017</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Spot Fleet target tracking</td>
<td>2016-11-15</td>
<td>You can set up target tracking scaling policies for your Spot Fleet. For more information, see <code>Scale Spot Fleet Using a Target Tracking Policy (p. 275)</code>.</td>
<td>17 November 2017</td>
</tr>
<tr>
<td>Spot Fleet integrates with Elastic Load Balancing</td>
<td>2016-11-15</td>
<td>You can attach one or more load balancers to a Spot Fleet.</td>
<td>10 November 2017</td>
</tr>
<tr>
<td>X1e instances</td>
<td>2016-11-15</td>
<td>X1e instances are ideally suited for high-performance databases, in-memory databases, and other memory-intensive enterprise applications. For more information, see <code>Memory Optimized Instances (p. 165)</code>.</td>
<td>28 November 2017</td>
</tr>
<tr>
<td>C5 instances</td>
<td>2016-11-15</td>
<td>C5 instances are designed for compute-heavy applications. For more information, see <code>Compute Optimized Instances (p. 161)</code>.</td>
<td>6 November 2017</td>
</tr>
<tr>
<td>Merge and split Convertible Reserved Instances</td>
<td>2016-11-15</td>
<td>You can exchange (merge) two or more Convertible Reserved Instances for a new Convertible Reserved Instance. You can also use the modification process to split a Convertible Reserved Instance into smaller reservations. For more information, see <code>Exchanging Convertible Reserved Instances (p. 224)</code>.</td>
<td>6 November 2017</td>
</tr>
<tr>
<td>P3 instances</td>
<td>2016-11-15</td>
<td>P3 instances are the next generation of compute-optimized GPU instances. For more information, see <code>Windows Accelerated Computing Instances (p. 178)</code>.</td>
<td>25 October 2017</td>
</tr>
<tr>
<td>Modify VPC tenancy</td>
<td>2016-11-15</td>
<td>You can change the instance tenancy attribute of a VPC from dedicated to default. For more information, see <code>Changing the Tenancy of a VPC (p. 317)</code>.</td>
<td>16 October 2017</td>
</tr>
<tr>
<td>Stop on interruption</td>
<td>2016-11-15</td>
<td>You can specify whether Amazon EC2 should stop or terminate Spot Instances when they are interrupted. For more information, see <code>Interruption Behavior (p. 286)</code>.</td>
<td>18 September 2017</td>
</tr>
<tr>
<td>Tag NAT gateways</td>
<td>2016-11-15</td>
<td>You can tag your NAT gateway. For more information, see <code>Tagging Your Resources (p. 966)</code>.</td>
<td>7 September 2017</td>
</tr>
<tr>
<td>Security group rule descriptions</td>
<td>2016-11-15</td>
<td>You can add descriptions to your security group rules. For more information, see <code>Security Group Rules (p. 584)</code>.</td>
<td>31 August 2017</td>
</tr>
<tr>
<td>Elastic Graphics</td>
<td>2016-11-15</td>
<td>Attach Elastic Graphics accelerators to your instances to accelerate the graphics performance of your applications. For more information, see <code>Amazon Elastic Graphics (p. 520)</code>.</td>
<td>29 August 2017</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Recover Elastic IP addresses</td>
<td>2016-11-15</td>
<td>If you release an Elastic IP address for use in a VPC, you might be able to recover it. For more information, see Recovering an Elastic IP Address (p. 678).</td>
<td>11 August 2017</td>
</tr>
<tr>
<td>Tag Spot Fleet instances</td>
<td>2016-11-15</td>
<td>You can configure your Spot Fleet to automatically tag the instances that it launches.</td>
<td>24 July 2017</td>
</tr>
<tr>
<td>G3 instances</td>
<td>2016-11-15</td>
<td>G3 instances provide a cost-effective, high-performance platform for graphics applications using DirectX or OpenGL. G3 instances also provide NVIDIA GRID Virtual Workstation features, supporting 4 monitors with resolutions up to 4096x2160. For more information, see Windows Accelerated Computing Instances (p. 178).</td>
<td>13 July 2017</td>
</tr>
<tr>
<td>Tag resources during creation</td>
<td>2016-11-15</td>
<td>You can apply tags to instances and volumes during creation. For more information, see Tagging Your Resources (p. 966). In addition, you can use tag-based resource-level permissions to control the tags that are applied. For more information see, Resource-Level Permissions for Tagging (p. 609).</td>
<td>28 March 2017</td>
</tr>
<tr>
<td>I3 instances</td>
<td>2016-11-15</td>
<td>I3 instances represent the next generation of storage optimized instances. For more information, see Storage Optimized Instances (p. 173).</td>
<td>23 February 2017</td>
</tr>
<tr>
<td>Perform modifications on attached EBS volumes</td>
<td>2016-11-15</td>
<td>With most EBS volumes attached to most EC2 instances, you can modify volume size, type, and IOPS without detaching the volume or stopping the instance. For more information, see Amazon EBS Elastic Volumes (p. 855).</td>
<td>13 February 2017</td>
</tr>
<tr>
<td>Attach an IAM role</td>
<td>2016-11-15</td>
<td>You can attach, detach, or replace an IAM role for an existing instance. For more information, see IAM Roles for Amazon EC2 (p. 646).</td>
<td>9 February 2017</td>
</tr>
<tr>
<td>Dedicated Spot Instances</td>
<td>2016-11-15</td>
<td>You can run Spot Instances on single-tenant hardware in a virtual private cloud (VPC). For more information, see Specifying a Tenancy for Your Spot Instances (p. 246).</td>
<td>19 January 2017</td>
</tr>
<tr>
<td>IPv6 support</td>
<td>2016-11-15</td>
<td>You can associate an IPv6 CIDR with your VPC and subnets, and assign IPv6 addresses to instances in your VPC. For more information, see Amazon EC2 Instance IP Addressing (p. 656).</td>
<td>1 December 2016</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>--------------------------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>R4 instances</td>
<td>2016-09-15</td>
<td>R4 instances represent the next generation of memory optimized instances. R4 instances are well-suited for memory-intensive, latency-sensitive workloads such as business intelligence (BI), data mining and analysis, in-memory databases, distributed web scale in-memory caching, and applications performance real-time processing of unstructured big data. For more information, see Memory Optimized Instances (p. 165)</td>
<td>30 November 2016</td>
</tr>
<tr>
<td>New t2.xlarge and t2.2xlarge instance types</td>
<td>2016-09-15</td>
<td>T2 instances are designed to provide moderate base performance and the capability to burst to significantly higher performance as required by your workload. They are intended for applications that need responsiveness, high performance for limited periods of time, and a low cost. For more information, see Burstable Performance Instances (p. 132).</td>
<td>30 November 2016</td>
</tr>
<tr>
<td>P2 instances</td>
<td>2016-09-15</td>
<td>P2 instances use NVIDIA Tesla K80 GPUs and are designed for general purpose GPU computing using the CUDA or OpenCL programming models. For more information, see Windows Accelerated Computing Instances (p. 178).</td>
<td>29 September 2016</td>
</tr>
<tr>
<td>m4.16xlarge instances</td>
<td>2016-04-01</td>
<td>Expands the range of the general-purpose M4 family with the introduction of m4.16xlarge instances, with 64 vCPUs and 256 GiB of RAM.</td>
<td>6 September 2016</td>
</tr>
<tr>
<td>Automatic scaling for Spot Fleet</td>
<td></td>
<td>You can now set up scaling policies for your Spot Fleet. For more information, see Automatic Scaling for Spot Fleet (p. 275).</td>
<td>1 September 2016</td>
</tr>
<tr>
<td>Elastic Network Adapter (ENA)</td>
<td>2016-04-01</td>
<td>You can now use ENA for enhanced networking. For more information, see Enhanced Networking Types (p. 700).</td>
<td>28 June 2016</td>
</tr>
<tr>
<td>Enhanced support for viewing and modifying longer IDs</td>
<td>2016-04-01</td>
<td>You can now view and modify longer ID settings for other IAM users, IAM roles, or the root user. For more information, see Resource IDs (p. 956).</td>
<td>23 June 2016</td>
</tr>
<tr>
<td>Copy encrypted Amazon EBS snapshots between AWS accounts</td>
<td>2016-04-01</td>
<td>You can now copy encrypted EBS snapshots between AWS accounts. For more information, see Copying an Amazon EBS Snapshot (p. 837).</td>
<td>21 June 2016</td>
</tr>
<tr>
<td>Capture a screenshot of an instance console</td>
<td>2015-10-01</td>
<td>You can now obtain additional information when debugging instances that are unreachable. For more information, see Troubleshoot an Unreachable Instance (p. 1051).</td>
<td>24 May 2016</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>X1 instances</td>
<td>2015-10-01</td>
<td>Memory-optimized instances designed for running in-memory databases, big data processing engines, and high performance computing (HPC) applications. For more information, see Memory Optimized Instances (p. 165).</td>
<td>18 May 2016</td>
</tr>
<tr>
<td>Two new EBS volume types</td>
<td>2015-10-01</td>
<td>You can now create Throughput Optimized HDD (st1) and Cold HDD (sc1) volumes. For more information, see Amazon EBS Volume Types (p. 786).</td>
<td>19 April 2016</td>
</tr>
<tr>
<td>Added new NetworkPacketsIn and NetworkPacketsOut metrics for Amazon EC2</td>
<td>2015-10-01</td>
<td>Added new NetworkPacketsIn and NetworkPacketsOut metrics for Amazon EC2. For more information, see Instance Metrics (p. 548).</td>
<td>23 March 2016</td>
</tr>
<tr>
<td>CloudWatch metrics for Spot Fleet</td>
<td>2015-10-01</td>
<td>You can now get CloudWatch metrics for your Spot Fleet. For more information, see CloudWatch Metrics for Spot Fleet (p. 272).</td>
<td>21 March 2016</td>
</tr>
<tr>
<td>Scheduled Instances</td>
<td>2015-10-01</td>
<td>Scheduled Reserved Instances (Scheduled Instances) enable you to purchase capacity reservations that recur on a daily, weekly, or monthly basis, with a specified start time and duration. For more information, see Scheduled Reserved Instances (p. 227).</td>
<td>13 January 2016</td>
</tr>
<tr>
<td>Longer resource IDs</td>
<td>2015-10-01</td>
<td>We're gradually introducing longer length IDs for some Amazon EC2 and Amazon EBS resource types. During the opt-in period, you can enable the longer ID format for supported resource types. For more information, see Resource IDs (p. 956).</td>
<td>13 January 2016</td>
</tr>
<tr>
<td>ClassicLink DNS support</td>
<td>2015-10-01</td>
<td>You can enable ClassicLink DNS support for your VPC so that DNS hostnames that are addressed between linked EC2-Classic instances and instances in the VPC resolve to private IP addresses and not public IP addresses. For more information, see Enabling ClassicLink DNS Support (p. 766).</td>
<td>11 January 2016</td>
</tr>
<tr>
<td>New t2.nano instance type</td>
<td>2015-10-01</td>
<td>T2 instances are designed to provide moderate base performance and the capability to burst to significantly higher performance as required by your workload. They are intended for applications that need responsiveness, high performance for limited periods of time, and a low cost. For more information, see Burstable Performance Instances (p. 132).</td>
<td>15 December 2015</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Dedicated hosts</td>
<td>2015-10-01</td>
<td>An Amazon EC2 Dedicated host is a physical server with instance capacity dedicated for your use. For more information, see Dedicated Hosts (p. 294).</td>
<td>23 November 2015</td>
</tr>
<tr>
<td>Spot Instance duration</td>
<td>2015-10-01</td>
<td>You can now specify a duration for your Spot Instances. For more information, see Specifying a Duration for Your Spot Instances (p. 246).</td>
<td>6 October 2015</td>
</tr>
<tr>
<td>Spot Fleet modify request</td>
<td>2015-10-01</td>
<td>You can now modify the target capacity of your Spot Fleet request. For more information, see Modifying a Spot Fleet Request (p. 262).</td>
<td>29 September 2015</td>
</tr>
<tr>
<td>Spot Fleet diversified allocation strategy</td>
<td>2015-04-15</td>
<td>You can now allocate Spot Instances in multiple Spot pools using a single Spot Fleet request. For more information, see Allocation Strategy for Spot Instances (p. 236).</td>
<td>15 September 2015</td>
</tr>
<tr>
<td>Spot Fleet instance weighting</td>
<td>2015-04-15</td>
<td>You can now define the capacity units that each instance type contributes to your application's performance, and adjust the amount you are willing to pay for Spot Instances for each Spot pool accordingly. For more information, see Spot Fleet Instance Weighting (p. 239).</td>
<td>31 August 2015</td>
</tr>
<tr>
<td>New reboot alarm action and new IAM role for use with alarm actions</td>
<td></td>
<td>Added the reboot alarm action and new IAM role for use with alarm actions. For more information, see Create Alarms That Stop, Terminate, Reboot, or Recover an Instance (p. 565).</td>
<td>23 July 2015</td>
</tr>
<tr>
<td>New t2.large instance type</td>
<td></td>
<td>T2 instances are designed to provide moderate base performance and the capability to burst to significantly higher performance as required by your workload. They are intended for applications that need responsiveness, high performance for limited periods of time, and a low cost. For more information, see Burstable Performance Instances (p. 132).</td>
<td>16 June 2015</td>
</tr>
<tr>
<td>M4 instances</td>
<td></td>
<td>The next generation of general-purpose instances that provide a balance of compute, memory, and network resources. M4 instances are powered by a custom Intel 2.4 GHz Intel® Xeon® E5 2676v3 (Haswell) processor with AVX2.</td>
<td>11 June 2015</td>
</tr>
<tr>
<td>Spot Fleets</td>
<td>2015-04-15</td>
<td>You can manage a collection, or fleet, of Spot Instances instead of managing separate Spot Instance requests. For more information, see How Spot Fleet Works (p. 235).</td>
<td>18 May 2015</td>
</tr>
<tr>
<td>Migrate Elastic IP addresses to EC2-Classic</td>
<td>2015-04-15</td>
<td>You can migrate an Elastic IP address that you’ve allocated for use in EC2-Classic to be used in a VPC. For more information, see Migrating an Elastic IP Address from EC2-Classic (p. 757).</td>
<td>15 May 2015</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Importing VMs with multiple disks as AMIs</td>
<td>2015-03-01</td>
<td>The VM Import process now supports importing VMs with multiple disks as AMIs. For more information, see <a href="Importing_VM.Import.Export_User_Guide">Importing a VM as an Image Using VM Import/Export</a> in the <em>VM Import/Export User Guide</em>.</td>
<td>23 April 2015</td>
</tr>
<tr>
<td>New g2.8xlarge instance type</td>
<td></td>
<td>The new g2.8xlarge instance is backed by four high-performance NVIDIA GPUs, making it well suited for GPU compute workloads including large scale rendering, transcoding, machine learning, and other server-side workloads that require massive parallel processing power.</td>
<td>7 April 2015</td>
</tr>
<tr>
<td>D2 instances</td>
<td></td>
<td>Next generation Amazon EC2 dense-storage instances that are optimized for applications requiring sequential access to large amount of data on direct attached instance storage. D2 instances are designed to offer best price/performance in the dense-storage family. Powered by 2.4 GHz Intel® Xeon® E5 2676v3 (Haswell) processors, D2 instances improve on HS1 instances by providing additional compute power, more memory, and Enhanced Networking. In addition, D2 instances are available in four instance sizes with 6TB, 12TB, 24TB, and 48TB storage options. For more information, see <a href="Storage_Optimized_Instances">Storage Optimized Instances</a> (p. 173).</td>
<td>24 March 2015</td>
</tr>
<tr>
<td>Systems Manager</td>
<td></td>
<td>Systems Manager enables you to configure and manage your EC2 instances.</td>
<td>17 February 2015</td>
</tr>
<tr>
<td>Systems Manager for Microsoft SCVMM 1.5</td>
<td></td>
<td>You can now use Systems Manager for Microsoft SCVMM to launch an instance and to import a VM from SCVMM to Amazon EC2. For more information, see <a href="Creating_an_EC2_Instance">Creating an EC2 Instance</a> (p. 983) and <a href="Importing_Your_Virtual_Machine">Importing Your Virtual Machine</a> (p. 987).</td>
<td>21 January 2015</td>
</tr>
<tr>
<td>Automatic recovery for EC2 instances</td>
<td></td>
<td>You can create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance and automatically recovers the instance if it becomes impaired due to an underlying hardware failure or a problem that requires AWS involvement to repair. A recovered instance is identical to the original instance, including the instance ID, IP addresses, and all instance metadata. For more information, see <a href="Recover_Your_Instance">Recover Your Instance</a> (p. 393).</td>
<td>12 January 2015</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>C4 instances</td>
<td></td>
<td>Next-generation compute-optimized instances that provide very high CPU performance at an economical price. C4 instances are based on custom 2.9 GHz Intel® Xeon® E5-2666 v3 (Haswell) processors. With additional Turbo boost, the processor clock speed in C4 instances can reach as high as 3.5Ghz with 1 or 2 core turbo. Expanding on the capabilities of C3 compute-optimized instances, C4 instances offer customers the highest processor performance among EC2 instances. These instances are ideally suited for high-traffic web applications, ad serving, batch processing, video encoding, distributed analytics, high-energy physics, genome analysis, and computational fluid dynamics. For more information, see <a href="https://aws.amazon.com/documentation/ec2/">Compute Optimized Instances (p. 161)</a>.</td>
<td>11 January 2015</td>
</tr>
<tr>
<td>ClassicLink</td>
<td>2014-10-01</td>
<td>ClassicLink enables you to link your EC2-Classic instance to a VPC in your account. You can associate VPC security groups with the EC2-Classic instance, enabling communication between your EC2-Classic instance and instances in your VPC using private IP addresses. For more information, see <a href="https://aws.amazon.com/documentation/ec2/">ClassicLink (p. 760)</a>.</td>
<td>7 January 2015</td>
</tr>
<tr>
<td>Spot Instance termination notices</td>
<td></td>
<td>The best way to protect against Spot Instance interruption is to architect your application to be fault tolerant. In addition, you can take advantage of Spot Instance termination notices, which provide a two-minute warning before Amazon EC2 must terminate your Spot Instance. For more information, see <a href="https://aws.amazon.com/documentation/ec2/">Spot Instance Interruption Notices (p. 289)</a>.</td>
<td>5 January 2015</td>
</tr>
<tr>
<td>Systems Manager for Microsoft SCVMM</td>
<td></td>
<td>Systems Manager for Microsoft SCVMM provides a simple, easy-to-use interface for managing AWS resources, such as EC2 instances, from Microsoft SCVMM. For more information, see <a href="https://aws.amazon.com/documentation/ec2/">AWS Systems Manager for Microsoft System Center VMM (p. 978)</a>.</td>
<td>29 October 2014</td>
</tr>
<tr>
<td>DescribeVolumes pagination support</td>
<td>2014-09-01</td>
<td>The DescribeVolumes API call now supports the pagination of results with the MaxResults and NextToken parameters. For more information, see <a href="https://aws.amazon.com/documentation/ec2/">DescribeVolumes in the Amazon EC2 API Reference</a>.</td>
<td>23 October 2014</td>
</tr>
</tbody>
</table>
### Feature

<table>
<thead>
<tr>
<th>Feature</th>
<th>API Version</th>
<th>Description</th>
<th>Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Added support for Amazon CloudWatch Logs</td>
<td></td>
<td>You can use Amazon CloudWatch Logs to monitor, store, and access your system, application, and custom log files from your instances or other sources. You can then retrieve the associated log data from CloudWatch Logs using the Amazon CloudWatch console, the CloudWatch Logs commands in the AWS CLI, or the CloudWatch Logs SDK.</td>
<td>10 July 2014</td>
</tr>
<tr>
<td>T2 instances</td>
<td>2014-06-15</td>
<td>T2 instances are designed to provide moderate base performance and the capability to burst to significantly higher performance as required by your workload. They are intended for applications that need responsiveness, high performance for limited periods of time, and a low cost. For more information, see Burstable Performance Instances (p. 132).</td>
<td>30 June 2014</td>
</tr>
<tr>
<td>New EC2 Service Limits page</td>
<td></td>
<td>Use the EC2 Service Limits page in the Amazon EC2 console to view the current limits for resources provided by Amazon EC2 and Amazon VPC, on a per-region basis.</td>
<td>19 June 2014</td>
</tr>
<tr>
<td>Amazon EBS General Purpose SSD Volumes</td>
<td>2014-05-01</td>
<td>General Purpose SSD volumes offer cost-effective storage that is ideal for a broad range of workloads. These volumes deliver single-digit millisecond latencies, the ability to burst to 3,000 IOPS for extended periods of time, and a base performance of 3 IOPS/GB. General Purpose SSD volumes can range in size from 1 GB to 1 TB. For more information, see General Purpose SSD (gp2) Volumes (p. 788).</td>
<td>16 June 2014</td>
</tr>
<tr>
<td>Windows Server 2012 R2</td>
<td></td>
<td>AMIs for Windows Server 2012 R2 use the new AWS PV drivers. For more information, see AWS PV Drivers (p. 432).</td>
<td>3 June 2014</td>
</tr>
<tr>
<td>AWS Management Pack</td>
<td></td>
<td>AWS Management Pack now supports for System Center Operations Manager 2012 R2. For more information, see AWS Management Pack for Microsoft System Center (p. 992).</td>
<td>22 May 2014</td>
</tr>
<tr>
<td>Amazon EBS encryption</td>
<td>2014-05-01</td>
<td>Amazon EBS encryption offers seamless encryption of EBS data volumes and snapshots, eliminating the need to build and maintain a secure key management infrastructure. EBS encryption enables data at rest security by encrypting your data using Amazon-managed keys. The encryption occurs on the servers that host EC2 instances, providing encryption of data as it moves between EC2 instances and EBS storage. For more information, see Amazon EBS Encryption (p. 865).</td>
<td>21 May 2014</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>R3 instances</td>
<td>2014-02-01</td>
<td>Next generation memory-optimized instances with the best price point per GiB of RAM and high performance. These instances are ideally suited for relational and NoSQL databases, in-memory analytics solutions, scientific computing, and other memory-intensive applications that can benefit from the high memory per vCPU, high compute performance, and enhanced networking capabilities of R3 instances. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.</td>
<td>9 April 2014</td>
</tr>
<tr>
<td>Amazon EC2 Usage Reports</td>
<td></td>
<td>Amazon EC2 Usage Reports is a set of reports that shows cost and usage data of your usage of EC2. For more information, see Amazon EC2 Usage Reports (p. 976).</td>
<td>28 January 2014</td>
</tr>
<tr>
<td>Additional M3 instances</td>
<td>2013-10-15</td>
<td>The M3 instance sizes m3.medium and m3.large are now supported. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.</td>
<td>20 January 2014</td>
</tr>
<tr>
<td>I2 instances</td>
<td>2013-10-15</td>
<td>These instances provide very high IOPS. I2 instances also support enhanced networking that delivers improve inter-instance latencies, lower network jitter, and significantly higher packet per second (PPS) performance. For more information, see Storage Optimized Instances (p. 173).</td>
<td>19 December 2013</td>
</tr>
<tr>
<td>Updated M3 instances</td>
<td>2013-10-15</td>
<td>The M3 instance sizes, m3.xlarge and m3.2xlarge now support instance store with SSD volumes.</td>
<td>19 December 2013</td>
</tr>
<tr>
<td>Resource-level permissions for RunInstances</td>
<td>2013-10-15</td>
<td>You can now create policies in AWS Identity and Access Management to control resource-level permissions for the Amazon EC2 RunInstances API action. For more information and example policies, see Controlling Access to Amazon EC2 Resources (p. 597).</td>
<td>20 November 2013</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>C3 instances</td>
<td>2013-10-15</td>
<td>Compute-optimized instances that provide very high CPU performance at an economical price. C3 instances also support enhanced networking that delivers improved inter-instance latencies, lower network jitter, and significantly higher packet per second (PPS) performance. These instances are ideally suited for high-traffic web applications, ad serving, batch processing, video encoding, distributed analytics, high-energy physics, genome analysis, and computational fluid dynamics. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.</td>
<td>14 November 2013</td>
</tr>
<tr>
<td>Launching an instance from the AWS Marketplace</td>
<td></td>
<td>You can now launch an instance from the AWS Marketplace using the Amazon EC2 launch wizard. For more information, see Launching an AWS Marketplace Instance (p. 350).</td>
<td>11 November 2013</td>
</tr>
<tr>
<td>G2 instances</td>
<td>2013-10-01</td>
<td>These instances are ideally suited for video creation services, 3D visualizations, streaming graphics-intensive applications, and other server-side workloads requiring massive parallel processing power. For more information, see Windows Accelerated Computing Instances (p. 178).</td>
<td>4 November 2013</td>
</tr>
<tr>
<td>New launch wizard</td>
<td></td>
<td>There is a new and redesigned EC2 launch wizard. For more information, see Launching an Instance Using the Launch Instance Wizard (p. 334).</td>
<td>10 October 2013</td>
</tr>
<tr>
<td>Modifying Amazon EC2 Reserved Instances</td>
<td>2013-08-15</td>
<td>You can now modify Reserved Instances in a region.</td>
<td>11 September 2013</td>
</tr>
<tr>
<td>Assigning a public IP address</td>
<td>2013-07-15</td>
<td>You can now assign a public IP address when you launch an instance in a VPC. For more information, see Assigning a Public IPv4 Address During Instance Launch (p. 661).</td>
<td>20 August 2013</td>
</tr>
<tr>
<td>Granting resource-level permissions</td>
<td>2013-06-15</td>
<td>Amazon EC2 supports new Amazon Resource Names (ARNs) and condition keys. For more information, see IAM Policies for Amazon EC2 (p. 599).</td>
<td>8 July 2013</td>
</tr>
<tr>
<td>Incremental Snapshot Copies</td>
<td>2013-02-01</td>
<td>You can now perform incremental snapshot copies. For more information, see Copying an Amazon EBS Snapshot (p. 837).</td>
<td>11 June 2013</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>AWS Management Pack</td>
<td></td>
<td>The AWS Management Pack links Amazon EC2 instances and the Windows or Linux operating systems running inside them. The AWS Management Pack is an extension to Microsoft System Center Operations Manager. For more information, see AWS Management Pack for Microsoft System Center (p. 992).</td>
<td>8 May 2013</td>
</tr>
<tr>
<td>New Tags page</td>
<td></td>
<td>There is a new Tags page in the Amazon EC2 console. For more information, see Tagging Your Amazon EC2 Resources (p. 964).</td>
<td>04 April 2013</td>
</tr>
<tr>
<td>Additional EBS-optimized instance types</td>
<td>2013-02-01</td>
<td>The following instance types can now be launched as EBS-optimized instances: c1.xlarge, m2.2xlarge, m3.xlarge, and m3.2xlarge. For more information, see Amazon EBS-Optimized Instances (p. 876).</td>
<td>19 March 2013</td>
</tr>
<tr>
<td>PV Drivers</td>
<td></td>
<td>To learn how to upgrade the paravirtualized (PV) drivers on your Windows AMI, see Upgrading PV Drivers On Your Windows Instances (p. 436).</td>
<td>March 2013</td>
</tr>
<tr>
<td>Copy an AMI from one region to another</td>
<td>2013-02-01</td>
<td>You can copy an AMI from one region to another, enabling you to launch consistent instances in more than one AWS region quickly and easily. For more information, see Copying an AMI (p. 72).</td>
<td>11 March 2013</td>
</tr>
<tr>
<td>Launch instances into a default VPC</td>
<td>2013-02-01</td>
<td>Your AWS account is capable of launching instances into either EC2-Classic or a VPC, or only into a VPC, on a region-by-region basis. If you can launch instances only into a VPC, we create a default VPC for you. When you launch an instance, we launch it into your default VPC, unless you create a nondefault VPC and specify it when you launch the instance.</td>
<td>11 March 2013</td>
</tr>
<tr>
<td>High-memory cluster (cr1.8xlarge) instance type</td>
<td>2012-12-01</td>
<td>Have large amounts of memory coupled with high CPU and network performance. These instances are well suited for in-memory analytics, graph analysis, and scientific computing applications.</td>
<td>21 January 2013</td>
</tr>
<tr>
<td>High storage (hs1.8xlarge) instance type</td>
<td>2012-12-01</td>
<td>High storage instances provide very high storage density and high sequential read and write performance per instance. They are well-suited for data warehousing, Hadoop/MapReduce, and parallel file systems.</td>
<td>20 December 2012</td>
</tr>
<tr>
<td>EBS snapshot copy</td>
<td>2012-12-01</td>
<td>You can use snapshot copies to create backups of data, to create new Amazon EBS volumes, or to create Amazon Machine Images (AMIs). For more information, see Copying an Amazon EBS Snapshot (p. 837).</td>
<td>17 December 2012</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Updated EBS metrics and status checks for Provisioned IOPS SSD volumes</td>
<td>2012-10-01</td>
<td>Updated the EBS metrics to include two new metrics for Provisioned IOPS SSD volumes. For more information, see Amazon CloudWatch Metrics for Amazon EBS (p. 903). Also added new status checks for Provisioned IOPS SSD volumes. For more information, see EBS Volume Status Checks (p. 809).</td>
<td>20 November 2012</td>
</tr>
</tbody>
</table>
| Support for Windows Server 2012 | | Amazon EC2 now provides you with several pre-configured Windows Server 2012 AMIs. These AMIs are immediately available for use in every region and for every 64-bit instance type. The AMIs support the following languages:  
• English  
• Chinese Simplified  
• Chinese Traditional  
• Chinese Traditional Hong Kong  
• Japanese  
• Korean  
• Portuguese  
• Portuguese Brazil  
• Czech  
• Dutch  
• French  
• German  
• Hungarian  
• Italian  
• Polish  
• Russian  
• Spanish  
• Swedish  
• Turkish | 19 November 2012 |
<p>| M3 instances | 2012-10-01 | There are new M3 extra-large and M3 double-extra-large instance types. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types. | 31 October 2012     |
| Spot Instance request status | 2012-10-01 | Spot Instance request status makes it easy to determine the state of your Spot requests.                                                                                                                   | 14 October 2012     |</p>
<table>
<thead>
<tr>
<th>Feature</th>
<th>API Version</th>
<th>Description</th>
<th>Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon EC2 Reserved Instance Marketplace</td>
<td>2012-08-15</td>
<td>The Reserved Instance Marketplace matches sellers who have Amazon EC2 Reserved Instances that they no longer need with buyers who are looking to purchase additional capacity. Reserved Instances bought and sold through the Reserved Instance Marketplace work like any other Reserved Instances, except that they can have less than a full standard term remaining and can be sold at different prices.</td>
<td>11 September 2012</td>
</tr>
<tr>
<td>Provisioned IOPS SSD for Amazon EBS</td>
<td>2012-07-20</td>
<td>Provisioned IOPS SSD volumes deliver predictable, high performance for I/O intensive workloads, such as database applications, that rely on consistent and fast response times. For more information, see Amazon EBS Volume Types (p. 786).</td>
<td>31 July 2012</td>
</tr>
<tr>
<td>High I/O instances for Amazon EC2</td>
<td>2012-06-15</td>
<td>High I/O instances provide very high, low latency, disk I/O performance using SSD-based local instance storage.</td>
<td>18 July 2012</td>
</tr>
</tbody>
</table>
| IAM roles on Amazon EC2 instances | 2012-06-01 | IAM roles for Amazon EC2 provide:  
- AWS access keys for applications running on Amazon EC2 instances.  
- Automatic rotation of the AWS access keys on the Amazon EC2 instance.  
- Granular permissions for applications running on Amazon EC2 instances that make requests to your AWS services. | 11 June 2012 |
| Spot Instance features that make it easier to get started and handle the potential of interruption. | | You can now manage your Spot Instances as follows:  
- Specify the amount you are willing to pay for Spot Instances using Auto Scaling launch configurations, and set up a schedule for specifying the amount you are willing to pay for Spot Instances. For more information, see Launching Spot Instances in Your Auto Scaling Group in the Amazon EC2 Auto Scaling User Guide.  
- Get notifications when instances are launched or terminated.  
- Use AWS CloudFormation templates to launch Spot Instances in a stack with AWS resources. | 7 June 2012 |
| EC2 instance export and timestamps for status checks for Amazon EC2 | 2012-05-01 | Added support for exporting Windows Server instances that you originally imported into EC2.  
Added support for timestamps on instance status and system status to indicate the date and time that a status check failed. | 25 May 2012 |
<table>
<thead>
<tr>
<th>Feature</th>
<th>API Version</th>
<th>Description</th>
<th>Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC2 instance export, and timestamps in instance and system status checks for Amazon VPC</td>
<td>2012-05-01</td>
<td>Added support for EC2 instance export to Citrix Xen, Microsoft Hyper-V, and VMware vSphere. Added support for timestamps in instance and system status checks.</td>
<td>25 May 2012</td>
</tr>
<tr>
<td>Cluster Compute Eight Extra Large instances</td>
<td>2012-04-01</td>
<td>Added support for cc2.8xlarge instances in a VPC.</td>
<td>26 April 2012</td>
</tr>
<tr>
<td>AWS Marketplace AMIs</td>
<td>2012-04-01</td>
<td>Added support for AWS Marketplace AMIs.</td>
<td>19 April 2012</td>
</tr>
<tr>
<td>Medium instances, support for 64-bit on all AMIs</td>
<td>2011-12-15</td>
<td>Added support for a new instance type and 64-bit information.</td>
<td>7 March 2012</td>
</tr>
<tr>
<td>Reserved Instance pricing tiers</td>
<td>2011-12-15</td>
<td>Added a new section discussing how to take advantage of the discount pricing that is built into the Reserved Instance pricing tiers.</td>
<td>5 March 2012</td>
</tr>
<tr>
<td>Elastic Network Interfaces (ENIs) for EC2 instances in Amazon Virtual Private Cloud</td>
<td>2011-12-01</td>
<td>Added new section about elastic network interfaces (ENIs) for EC2 instances in a VPC. For more information, see Elastic Network Interfaces (p. 679).</td>
<td>21 December 2011</td>
</tr>
<tr>
<td>New offering types for Amazon EC2 Reserved Instances</td>
<td>2011-11-01</td>
<td>You can choose from a variety of Reserved Instance offerings that address your projected use of the instance.</td>
<td>01 December 2011</td>
</tr>
<tr>
<td>Amazon EC2 instance status</td>
<td>2011-11-01</td>
<td>You can view additional details about the status of your instances, including scheduled events planned by AWS that might have an impact on your instances. These operational activities include instance reboots required to apply software updates or security patches, or instance retirements required where there are hardware issues. For more information, see Monitoring the Status of Your Instances (p. 535).</td>
<td>16 November 2011</td>
</tr>
<tr>
<td>Amazon EC2 Cluster Compute Instance Type</td>
<td></td>
<td>Added support for Cluster Compute Eight Extra Large (cc2.8xlarge) to Amazon EC2.</td>
<td>14 November 2011</td>
</tr>
<tr>
<td>Spot Instances in Amazon VPC</td>
<td>2011-07-15</td>
<td>Added information about the support for Spot Instances in Amazon VPC. With this update, users can launch Spot Instances a virtual private cloud (VPC). By launching Spot Instances in a VPC, users of Spot Instances can enjoy the benefits of Amazon VPC.</td>
<td>11 October 2011</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Simplified VM import process for users of the CLI tools</td>
<td>2011-07-15</td>
<td>The VM Import process is simplified with the enhanced functionality of ImportInstance and ImportVolume, which now will perform the upload of the images into Amazon EC2 after creating the import task. In addition, with the introduction of ResumeImport, users can restart an incomplete upload at the point the task stopped.</td>
<td>15 September 2011</td>
</tr>
<tr>
<td>Support for importing in VHD file format</td>
<td></td>
<td>VM Import can now import virtual machine image files in VHD format. The VHD file format is compatible with the Citrix Xen and Microsoft Hyper-V virtualization platforms. With this release, VM Import now supports RAW, VHD and VMDK (VMware ESX-compatible) image formats. For more information, see the VM Import/Export User Guide.</td>
<td>24 August 2011</td>
</tr>
<tr>
<td>Support for Windows Server 2003 R2</td>
<td></td>
<td>VM Import now supports Windows Server 2003 (R2). With this release, VM Import supports all versions of Windows Server supported by Amazon EC2.</td>
<td>24 August 2011</td>
</tr>
<tr>
<td>Update to the Amazon EC2 VM Import Connector for VMware vCenter</td>
<td></td>
<td>Added information about the 1.1 version of the Amazon EC2 VM Import Connector for VMware vCenter virtual appliance (Connector). This update includes proxy support for Internet access, better error handling, improved task progress bar accuracy, and several bug fixes.</td>
<td>27 June 2011</td>
</tr>
<tr>
<td>Spot Instances Availability Zone pricing changes</td>
<td>2011-05-15</td>
<td>Added information about the Spot Instances Availability Zone pricing feature. In this release, we've added new Availability Zone pricing options as part of the information returned when you query for Spot Instance requests and Spot price history. These additions make it easier to determine the price required to launch a Spot Instance into a particular Availability Zone.</td>
<td>26 May 2011</td>
</tr>
<tr>
<td>AWS Identity and Access Management</td>
<td></td>
<td>Added information about AWS Identity and Access Management (IAM), which enables users to specify which Amazon EC2 actions a user can use with Amazon EC2 resources in general. For more information, see Controlling Access to Amazon EC2 Resources (p. 597).</td>
<td>26 April 2011</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Dedicated instances</td>
<td></td>
<td>Launched within your Amazon Virtual Private Cloud (Amazon VPC), Dedicated Instances are instances that are physically isolated at the host hardware level. Dedicated Instances let you take advantage of Amazon VPC and the AWS cloud, with benefits including on-demand elastic provisioning and pay only for what you use, while isolating your Amazon EC2 compute instances at the hardware level. For more information, see Dedicated Instances (p. 312).</td>
<td>27 March 2011</td>
</tr>
<tr>
<td>Reserved Instances updates to the AWS Management Console</td>
<td></td>
<td>Updates to the AWS Management Console make it easier for users to view their Reserved Instances and purchase additional Reserved Instances, including Dedicated Reserved Instances.</td>
<td>27 March 2011</td>
</tr>
<tr>
<td>Support for Windows Server 2008 R2</td>
<td></td>
<td>Amazon EC2 now provides you with several pre-configured Windows Server 2008 R2 AMIs. These AMIs are immediately available for use in every region and in most 64-bit instance types, excluding t1.micro and HPC families. The AMIs will support multiple languages.</td>
<td>15 March 2011</td>
</tr>
<tr>
<td>Metadata information</td>
<td>2011-01-01</td>
<td>Added information about metadata to reflect changes in the 2011-01-01 release. For more information, see Instance Metadata and User Data (p. 478) and Instance Metadata Categories (p. 482).</td>
<td>11 March 2011</td>
</tr>
<tr>
<td>Amazon EC2 VM Import Connector for VMware vCenter</td>
<td></td>
<td>Added information about the Amazon EC2 VM Import Connector for VMware vCenter virtual appliance (Connector). The Connector is a plug-in for VMware vCenter that integrates with VMware vSphere Client and provides a graphical user interface that you can use to import your VMware virtual machines to Amazon EC2.</td>
<td>3 March 2011</td>
</tr>
<tr>
<td>Force volume detachment</td>
<td></td>
<td>You can now use the AWS Management Console to force the detachment of an Amazon EBS volume from an instance. For more information, see Detaching an Amazon EBS Volume from an Instance (p. 816).</td>
<td>23 February 2011</td>
</tr>
<tr>
<td>Instance termination protection</td>
<td></td>
<td>You can now use the AWS Management Console to prevent an instance from being terminated. For more information, see Enabling Termination Protection for an Instance (p. 390).</td>
<td>23 February 2011</td>
</tr>
<tr>
<td>VM Import</td>
<td>2010-11-15</td>
<td>Added information about VM Import, which allows you to import a virtual machine or volume into Amazon EC2. For more information, see the VM Import/Export User Guide.</td>
<td>15 December 2010</td>
</tr>
<tr>
<td>Feature</td>
<td>API Version</td>
<td>Description</td>
<td>Release Date</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Basic monitoring for instances</td>
<td>2010-08-31</td>
<td>Added information about basic monitoring for EC2 instances.</td>
<td>12 December 2010</td>
</tr>
<tr>
<td>Filters and Tags</td>
<td>2010-08-31</td>
<td>Added information about listing, filtering, and tagging resources. For more information, see Listing and Filtering Your Resources (p. 961) and Tagging Your Amazon EC2 Resources (p. 964).</td>
<td>19 September 2010</td>
</tr>
<tr>
<td>Idempotent Instance Launch</td>
<td>2010-08-31</td>
<td>Added information about ensuring idempotency when running instances.</td>
<td>19 September 2010</td>
</tr>
<tr>
<td>Micro instances</td>
<td>2010-06-15</td>
<td>Amazon EC2 offers the t1.micro instance type for certain types of applications. For more information, see Burstable Performance Instances (p. 132).</td>
<td>8 September 2010</td>
</tr>
<tr>
<td>AWS Identity and Access Management for Amazon EC2</td>
<td></td>
<td>Amazon EC2 now integrates with AWS Identity and Access Management (IAM). For more information, see Controlling Access to Amazon EC2 Resources (p. 597).</td>
<td>2 September 2010</td>
</tr>
<tr>
<td>Cluster instances</td>
<td>2010-06-15</td>
<td>Amazon EC2 offers cluster compute instances for high-performance computing (HPC) applications. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.</td>
<td>12 July 2010</td>
</tr>
<tr>
<td>Amazon VPC IP Address Designation</td>
<td>2010-06-15</td>
<td>Amazon VPC users can now specify the IP address to assign an instance launched in a VPC.</td>
<td>12 July 2010</td>
</tr>
<tr>
<td>Amazon CloudWatch Monitoring for Amazon EBS Volumes</td>
<td></td>
<td>Amazon CloudWatch monitoring is now automatically available for Amazon EBS volumes. For more information, see Amazon CloudWatch Metrics for Amazon EBS (p. 903).</td>
<td>14 June 2010</td>
</tr>
<tr>
<td>High-memory extra large instances</td>
<td>2009-11-30</td>
<td>Amazon EC2 now supports a High-Memory Extra Large (m2.xlarge) instance type. For more information about the hardware specifications for each Amazon EC2 instance type, see Amazon EC2 Instance Types.</td>
<td>22 February 2010</td>
</tr>
<tr>
<td>Reserved Instances with Windows</td>
<td></td>
<td>Amazon EC2 now supports Reserved Instances with Windows.</td>
<td>22 February 2010</td>
</tr>
</tbody>
</table>
AWS Glossary

For the latest AWS terminology, see the AWS Glossary in the AWS General Reference.