



AWS Whitepaper

Amazon EC2 Reserved Instances and Other AWS Reservation Models



Amazon EC2 Reserved Instances and Other AWS Reservation Models: AWS Whitepaper

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Abstract	1
Abstract	1
Introduction	2
Amazon EC2 Reserved Instances	3
Reserved Instances payment options	3
Standard vs. Convertible offering classes	4
Regional and zonal Reserved Instances	5
Differences between regional and zonal Reserved Instances	5
Limitations for instance size flexibility	6
Maximizing Utilization with Size Flexibility in Regional Reserved Instances	6
Normalization factor for dedicated EC2 instances	8
Normalization factor for bare metal instances	9
Savings Plans	11
Reservation models for other AWS services	13
Amazon RDS reserved DB instances	13
Amazon ElastiCache reserved nodes	13
Amazon OpenSearch Service Reserved Instances	14
Amazon Redshift reserved nodes	14
Amazon DynamoDB reservations	14
Reserved Instances billing	16
Usage billing	16
Consolidated billing	17
Reserved Instances: Capacity reservations	18
Blended rates	18
How discounts are applied	18
Maximizing the value of reservations	20
Measure success	20
Maximize discounts by standardizing instance type	20
Reservation management techniques	21
Reserved Instance Marketplace	22
AWS Cost Explorer	22
AWS Cost and Usage Report	23
Reserved Instances on your cost and usage report	23
AWS Trusted Advisor	24

Conclusion	25
Contributors	26
Document revisions	27
Notices	28

Amazon EC2 Reserved Instances and Other AWS Reservation Models

Publication date: March 29, 2021 ([Document revisions](#))

Abstract

This document is part of a series of AWS whitepapers designed to support your cloud journey, and discusses Amazon EC2 Reserved Instances and reservation models for other AWS services. Its aim is to empower you to maximize the value of your investments, improve forecasting accuracy and cost predictability, create a culture of ownership and cost transparency, and continuously measure your optimization status.

Introduction

The cloud is well suited for variable workloads and rapid deployment, yet many cloud-based workloads follow a more predictable pattern. For such applications, your organization can achieve significant cost savings by using [Amazon Elastic Compute Cloud \(Amazon EC2\) Reserved Instances](#). Amazon EC2 Reserved Instances enable your organization to commit to usage parameters at the time of purchase to achieve a lower hourly rate. Reservation models are also available for [Amazon Relational Database Service \(Amazon RDS\)](#), [Amazon ElastiCache](#), [Amazon OpenSearch Service \(OpenSearch Service\)](#), [Amazon Redshift](#), and [Amazon DynamoDB](#). This whitepaper discusses Amazon EC2 Reserved Instances and the reservation models for these other AWS services.

Amazon EC2 Reserved Instances

When you purchase Reserved Instances, you make a one-year or three-year commitment and receive a billing discount of up to 72 percent in return. When used for the appropriate workloads, Reserved Instances can save you a lot of money.

Note that a Reserved Instance is not an instance dedicated to your organization. It is a billing discount applied to the use of On-Demand Instances in your account. These On-Demand Instances must match certain attributes of the Reserved Instances you purchased to benefit from the billing discount. You pay for the entire term of a Reserved Instance, regardless of actual usage, so your cost savings are closely tied to use. Therefore, it is important to plan and monitor your usage to make the most of your investment.

When you purchase a Reserved Instance in a specific Availability Zone, it provides a capacity reservation. This improves the likelihood that the compute capacity you need is available in a specific Availability Zone when you need it. A Reserved Instance purchased for an AWS Region does not provide capacity reservation.

Reserved Instances payment options

You can purchase Reserved Instances through the AWS Management Console. The following payment options are available for most Reserved Instances:

- **No Upfront** – No upfront payment is required. You are billed a discounted hourly rate for every hour within the term, regardless of whether the Reserved Instance is being used. No Upfront Reserved Instances are based on a contractual obligation to pay monthly for the entire term of the reservation. A successful billing history is required before you can purchase No Upfront Reserved Instances.
- **Partial Upfront** – A portion of the cost must be paid up front and the remaining hours in the term are billed at a discounted hourly rate, regardless of whether you're using the Reserved Instance.
- **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.

Reserved Instances with a higher upfront payment provide greater discounts. You can also find Reserved Instances offered by third-party sellers at lower prices and shorter terms on the [Reserved Instance Marketplace](#).

As you purchase more Reserved Instances, volume discounts begin to apply that let you save even more.

For more information, see [Amazon EC2 Reserved Instance Pricing](#).

Standard vs. Convertible offering classes

When you purchase a Reserved Instance, you can choose between a Standard or Convertible offering class.

Table 1 – Comparison of standard and Convertible Reserved Instances

Standard Reserved Instance	Convertible Reserved Instance
One-year to three-year term	One-year to three-year term
Enables you to <i>modify</i> Availability Zone, scope, networking type, and instance size (within the same instance type) of your Reserved Instance. For more information, see Modifying Reserved Instances .	Enables you to <i>exchange</i> 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. For more information, see Exchanging Convertible Reserved Instances .
Can be sold in the Reserved Instance Marketplace.	Cannot be sold in the Reserved Instance Marketplace.

Standard Reserved Instances typically provide the highest discount levels. One-year Standard Reserved Instances provide a similar discount to three-year Convertible Reserved Instances.

If you want to purchase capacity reservations, see [On-Demand Capacity Reservations](#).

Convertible Reserved Instances are useful when:

- Purchasing Reserved Instances in the payer account instead of a subaccount. You can more easily modify Convertible Reserved Instances to meet changing needs across your organization.
- Workloads are likely to change. In this case, a Convertible Reserved Instance enables you to adapt as needs evolve while still obtaining discounts and capacity reservations.
- You want to hedge against possible future price drops.
- You can't or don't want to ask teams to do capacity planning or forecasting.
- You expect compute usage to remain at the committed amount over the commitment period.

Regional and zonal Reserved Instances

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 2 – Comparison of regional and zonal Reserved Instances

	Regional Reserved Instances	Zonal Reserved Instances
Availability Zone flexibility	The Reserved Instance discount applies to instance usage in any Availability Zone in the specified Region.	No Availability Zone flexibility—the Reserved Instance discount applies to instance usage in the specified Availability Zone only.
Capacity reservation	No capacity reservation—a regional Reserved Instance	A zonal Reserved Instance provides a capacity reservati

	Regional Reserved Instances	Zonal Reserved Instances
	does <i>not</i> provide a capacity reservation.	are located 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 .	No instance size flexibility—the Reserved Instance discount applies to instance usage for the specified instance type and size only.

Limitations for instance size flexibility

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 Server, Windows Server with SQL Standard, Windows Server with SQL Server Enterprise, Windows Server with SQL Server Web, RHEL, and SUSE Linux Enterprise Server
- Reserved Instances for G4ad, G4dn, G5, and G5g instances.

Maximizing Utilization with Size Flexibility in Regional Reserved Instances

For additional flexibility, all Regional Linux Reserved Instances with shared tenancy apply to all sizes of instances within an instance family and an AWS Region, even if you are using them across multiple accounts via [Consolidated Billing](#). The only attributes that must be matched are the instance type (for example, m4), tenancy (must be default), and platform (must be Linux). All new

and existing Reserved Instances are sized according to a normalization factor based on instance size, as follows.

Table 3 – Regional Reserved Instance sizes and normalization factors

Instance size	Normalization factor
nano	0.25
micro	0.5
small	1
medium	2
large	4
xlarge	8
2xlarge	16
4xlarge	32
8xlarge	64
9xlarge	72
10xlarge	80
12xlarge	96
16xlarge	128
24xlarge	192
32xlarge	256

For example, if you have a Reserved Instance for a c4.8xlarge, it applies to any usage of a Linux c4 instance with shared tenancy in the AWS Region, such as:

- One c4.8xlarge instance

- Two c4.4xlarge instances
- Four c4.2xlarge instances
- Sixteen c4.large instances

It also includes combinations of instances, 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) Region and you have two running t2.small instances in your account in that Region, the billing benefit is applied in full to both instances.

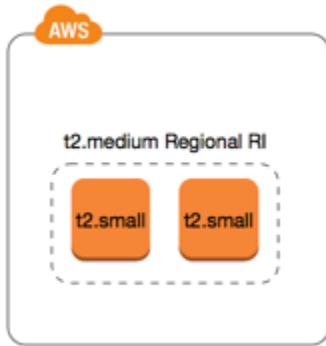


Figure 1 – Two t2.medium instances running in a Region

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.

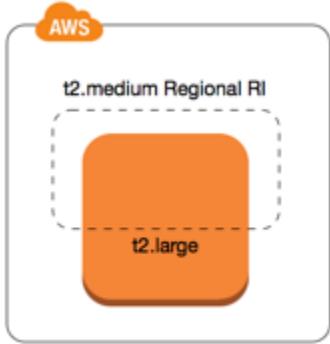


Figure 2 – One t2.large instance running in a Region

The normalization factor is also applied when modifying Reserved Instances.

Normalization factor for dedicated EC2 instances

For size inflexible RIs, the normalization factor is always 1. The normalization factor doesn't apply to EC2 instances that do not have size flexibility. The sole purpose of the normalization factor is

to provide an ability to match various EC2 instances to each other within a family so that you can exchange one type for another type. We do not support this use case for EC2 instances without size flexibility, hence normalization factor is not used and to keep our data model uniform across different EC2 use cases, we assign it an equivalent value of 1.

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.

The .metal instance sizes do not have a single normalization factor. They vary based on the specific instance family. For the most up-to-date list, see [Amazon EC2 Instance Types](#).

Table 4 – Bare metal instance sizes and normalization factors

Instance size	Normalization factor
a1.metal	32
c5.metal	192
c5d.metal	192
c5n.metal	144
c6g.metal	128
c6gd.metal	128
g4dn.metal	192
i3.metal	128
i3en.metal	192

Instance size	Normalization factor
m5.metal	192
m5d.metal	192
m5dn.metal	192
m5n.metal	192
m5zn.metal	96
m6g.metal	128
m6gd.metal	128
r5.metal	192
r5b.metal	192
r5d.metal	192
r5dn.metal	192
r5n.metal	192
r6g.metal	128
r6gd.metal	128
x2gd.metal	128
u-* .metal	896
z1d.metal	96

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) Region, 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) Region, and you have one running i3.metal instance in that Region, the billing benefit is applied in full to the i3.metal instance.

Savings Plans

Savings Plans is another flexible pricing model that provides savings of up to 72% on your AWS compute usage. This pricing model offers lower prices on Amazon EC2 instances usage, regardless of instance family, size, OS, tenancy or AWS Region, and also applies to AWS Fargate and AWS Lambda usage.

Savings Plans offer significant savings over On-Demand Instances, just like EC2 Reserved Instances, in exchange for a commitment to use a specific amount of compute power (measured in \$/hour) for a one or three-year period. You can sign up for Savings Plans for a one- or three-year term and easily manage your plans by taking advantage of recommendations, performance reporting and budget alerts in the AWS Cost Explorer.

AWS offers two types of Savings Plans:

- **Compute Savings Plans** provide the most flexibility and help to reduce your costs by up to 66% (just like Convertible RIs). These plans automatically apply to EC2 instance usage regardless of instance family, size, AZ, Region, operating system, or tenancy, and also apply to Fargate and Lambda usage. For example, with Compute Savings Plans, you can change from C4 to M5 instances, shift a workload from EU (Ireland) to Europe (London), or move a workload from Amazon EC2 to Fargate or Lambda at any time and automatically continue to pay the Savings Plans price.
- **EC2 Instance Savings Plans** provide the lowest prices, offering savings up to 72% (just like Standard RIs) in exchange for commitment to usage of individual instance families in a Region (for example, M5 usage in N. Virginia). This automatically reduces your cost on the selected

instance family in that region regardless of AZ, size, operating system, or tenancy. EC2 Instance Savings Plans give you the flexibility to change your usage between instances within a family in that Region. For example, you can move from c5.xlarge running Windows to c5.2xlarge running Linux and automatically benefit from the Savings Plans prices.

Note that Savings Plans does not provide a capacity reservation. You can however reserve capacity with On Demand Capacity Reservations and pay lower prices on them with Savings Plans.

You can continue purchasing RIs to maintain compatibility with your existing cost management processes, and your RIs will work along-side Savings Plans to reduce your overall bill. However, as your RIs expire we encourage you to sign up for Savings Plans as they offer the same savings as RIs, but with additional flexibility.

Reservation models for other AWS services

In addition to Amazon EC2, reservation models are available for Amazon RDS, Amazon ElastiCache, OpenSearch Service, Amazon Redshift, and Amazon DynamoDB.

Topics

- [Amazon RDS reserved DB instances](#)
- [Amazon ElastiCache reserved nodes](#)
- [Amazon OpenSearch Service Reserved Instances](#)
- [Amazon Redshift reserved nodes](#)
- [Amazon DynamoDB reservations](#)

Amazon RDS reserved DB instances

Similar to Amazon EC2 Reserved Instances, there are three payment options for Amazon RDS reserved DB instances: No Upfront, Partial Upfront, and All Upfront. All reserved DB instance types are available for Aurora, MySQL, MariaDB, PostgreSQL, Oracle, and SQL Server database engines.

Size-flexible reserved DB instances are available for Amazon Aurora, MariaDB, MySQL, PostgreSQL, and the “Bring Your Own License” (BYOL) edition of the Oracle database engine.

For more information about Amazon RDS reserved DB instances, see the following:

- [Amazon RDS Reserved Instances](#)
- [Working with Reserved DB Instances](#)
- [Amazon DynamoDB Pricing](#)

Amazon ElastiCache reserved nodes

Amazon ElastiCache reserved nodes give you the option to make a low, one-time payment for each cache node you want to reserve. In turn, you receive a significant discount on the hourly charge for that node. Amazon ElastiCache provides three reserved cache node types (Light Utilization, Medium Utilization, and Heavy Utilization) that enable you to balance the amount you pay up front with your effective hourly price. Based on your application workload and the amount of time you

plan to run them, Amazon ElastiCache Reserved Nodes might provide substantial savings over running on-demand Nodes. Reserved Cache Nodes are available for both Redis and Memcached.

For more information, see [Amazon ElastiCache reserved nodes](#).

Amazon OpenSearch Service Reserved Instances

Amazon OpenSearch Service (OpenSearch Service) Reserved Instances (RIs) offer significant discounts compared to standard On-Demand Instances. The instances themselves are identical—RIs are just a billing discount applied to On-Demand Instances in your account. For long-lived applications with predictable usage, RIs can provide considerable savings over time. OpenSearch Service RIs require one- or three-year terms and have three payment options that affect the discount rate.

For more information, see [Amazon OpenSearch Service Reserved Instances](#)

Amazon Redshift reserved nodes

In AWS, the charges that you accrue for using Amazon Redshift are based on compute nodes. Each compute node is billed at an hourly rate. The hourly rate varies depending on factors such as AWS Region, node type, and whether the node receives on-demand node pricing or reserved node pricing.

If you intend to keep an Amazon Redshift cluster running continuously for a prolonged period, you should consider purchasing reserved-node offerings. These offerings provide significant savings over on-demand pricing. However, they require you to reserve compute nodes and commit to paying for those nodes for either a one-year or a three-year duration.

For more information about Amazon Redshift reserved node pricing, see [Reserved Instance Pricing](#) and [Purchasing Amazon Redshift Reserved Nodes](#).

Amazon DynamoDB reservations

If you can predict your need for Amazon DynamoDB read-and-write throughput, reserved capacity offers significant savings over the normal price of DynamoDB provisioned throughput capacity. You pay a one-time upfront fee and commit to paying for a minimum usage level at specific hourly rates for the duration of the reserved capacity term. Any throughput you provision in excess of your reserved capacity is billed at standard rates for provisioned throughput.

Provisioned capacity mode might be best if you

- Have predictable application traffic.
- Run applications whose traffic is consistent or ramps gradually.
- Can forecast capacity requirements to control costs.

For more information, see [Pricing for Provisioned Capacity](#).

Reserved Instances billing

All Reserved Instances provide you with a discount compared to On-Demand Instance 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](#) specified for the Reserved Instance.

When Reserved Instances expire, you are charged On-Demand Instance rates. 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](#).

You can set up a billing alert to warn you when your bill exceeds a threshold that you define. For more information, see Monitoring Charges with Alerts and Notifications.

Usage billing

Except for DynamoDB reservations, which are billed based on throughput, reservations are billed for every clock-hour during the term you select, regardless of whether an instance is running or not. A clock-hour is defined as the standard 24-hour clock that runs from midnight to midnight and is divided into 24 hours (for example, 1:00:00 to 1:59:59 is one clock-hour).

A Reserved Instance billing benefit can be applied to a running instance on a per-second basis. Per-second billing is available for instances using an open-source Linux distribution, such as Amazon Linux and Ubuntu. Per-hour billing is used for commercial Linux distributions, such as Red Hat Enterprise Linux and SUSE Linux Enterprise Server.

A Reserved Instance billing benefit can apply to a maximum of 3600 seconds (one hour) of instance usage per clock-hour. You can run multiple instances concurrently, but can only receive the benefit of the Reserved Instance discount for a total of 3600 seconds per clock-hour. Instance usage that exceeds 3600 seconds in a clock-hour is billed at the On-Demand Instance rate.

For example, if you purchase one m4.xlarge Reserved Instance and run four m4.xlarge instances concurrently for one hour, one instance is charged at one hour of Reserved Instance usage and the other three instances are charged at three hours of On-Demand Instance usage.

However, if you purchase one m4.xlarge Reserved Instance and run four m4.xlarge instances for 15 minutes (900 seconds) each within the same hour, the total running time for the instances is one

hour, which results in one hour of Reserved Instance usage and 0 hours of On-Demand Instance usage.

	1:00	1:15	1:30	1:45
Instance 1				
Instance 2				
Instance 3				
Instance 4				

Figure 3 – Running four instances for 15 minutes each in the same hour

If multiple eligible instances are running concurrently, the Reserved Instance billing benefit is applied to all the instances at the same time up to a maximum of 3600 seconds in a clock-hour. Thereafter, the On-Demand Instance rates apply.

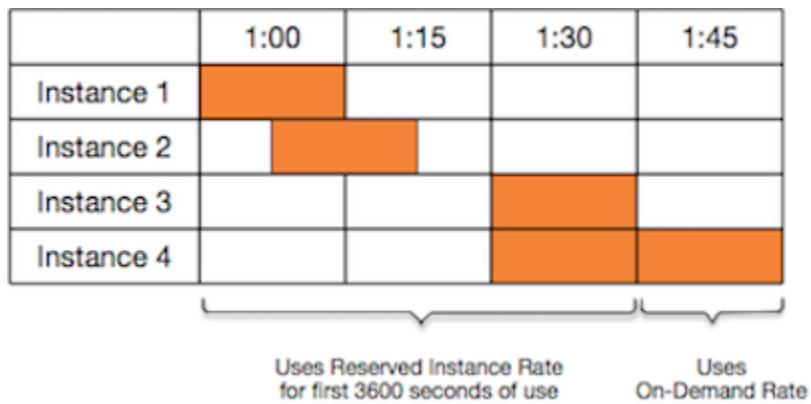


Figure 4 – Running four instances concurrently over the hour

You can find out about the charges and fees to your account by viewing the [AWS Billing and Cost Management](#) console. You can also examine your utilization and coverage, and receive reservation purchase recommendations, via [AWS Cost Explorer](#). You can dive deeper into your reservations and Reserved Instance discount allocation via the [AWS Cost and Usage Report](#).

For more information on Reserved Instance usage billing, see [Usage Billing](#).

Consolidated billing

AWS Organizations is an account management service that lets you consolidate multiple AWS accounts into an *organization* that you create and centrally manage. AWS Organizations includes consolidated billing and account management capabilities that enable you to better meet the

budgetary, security, and compliance needs of your business. For more information, see [What Is AWS Organizations?](#)

For more information on consolidated bills and how they are calculated, see [Understanding Consolidated Bills](#).

The pricing benefits of Reserved Instances are shared when the purchasing account is billed under a consolidated billing payer account. The instance usage across all member accounts is aggregated in the payer account every month. This is useful for companies that have different functional teams or groups, then, the normal Reserved Instance logic is applied to calculate the bill.

Reserved Instances: Capacity reservations

AWS also offers discounted hourly rates in exchange for an upfront fee and term contract. Services such as Amazon EC2 and Amazon RDS use this approach to sell reserved capacity for hourly use of Reserved Instances. For more information, see Reserved Instances in the *Amazon EC2 User Guide for Linux Instances* and Working with Reserved DB Instances in the *Amazon Relational Database Service User Guide*.

When you reserve capacity with Reserved Instances, your hourly usage is calculated at a discounted rate for instances of the same usage type in the same Availability Zone (AZ). When you launch additional instances of the same instance type in the same Availability Zone and exceed the number of instances in your reservation, AWS averages the rates of the Reserved Instances and the On-Demand Instances to give you a blended rate.

Blended rates

A line item for the blended rate of that instance is displayed on the bill of any member account that is running an instance that matches the specifications of a reservation in the organization.

The payer account of an organization can turn off Reserved Instance sharing for member accounts in that organization via the AWS Billing Preferences. This means that Reserved Instances are not shared between that member account and other member accounts. Each estimated bill is computed using the most recent set of preferences. For information on how to configure sharing, see [Turning Off Reserved Instance Sharing](#).

How discounts are applied

The application of Amazon EC2 Reserved Instances is based on instance attributes, including the following:

- **Instance type** – Instance types comprise varying combinations of CPU, memory, storage, and networking capacity (for example, m4.xlarge). This gives you the flexibility to choose the appropriate mix of resources for your applications, such as compute-optimized, storage-optimized, and so on. Each instance type includes one or more instance sizes, enabling you to scale your resources to the requirements of your target workload.
- **Platform** – You can purchase Reserved Instances for Amazon EC2 instances running Linux, Unix, SUSE Linux, Red Hat Enterprise Linux, Windows Server, and Microsoft SQL Server platforms.
- **Tenancy** – Reserved Instances can be default tenancy or dedicated tenancy.
- **Regional or zonal** – See [Regional and zonal Reserved Instances](#).

If you purchase a Reserved Instance and you already have a running instance that matches the attributes of the Reserved Instance, the billing benefit is immediately applied. You don't 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](#).

Maximizing the value of reservations

This section discusses how you can maximize the value of your reservations.

Topics

- [Measure success](#)
- [Maximize discounts by standardizing instance type](#)
- [Reservation management techniques](#)
- [Reserved Instance Marketplace](#)
- [AWS Cost Explorer](#)
- [AWS Cost and Usage Report](#)
- [AWS Trusted Advisor](#)

Measure success

Making the most of reservations means measuring your reservation coverage (portion of instances enjoying reservation discount benefits) and reservation utilization (degree to which purchased Reserved Instances are used). Establish a standardized review cadence in which you focus on the following questions:

- Do you need to modify any of our existing reservations to increase utilization?
- Are any currently utilized reservations expiring?
- Do you need to purchase any reservations to increase your coverage?

A standardized review cadence ensures that issues are surfaced and addressed in a timely manner. As your RIs expire we encourage you to sign up for Savings Plans as they offer the same savings as RIs, but with additional flexibility.

Maximize discounts by standardizing instance type

By standardizing the instance types that your organization uses, you can ensure that deployments match the characteristics of your reservations to maximize your discounts. Standardization maximizes utilization and minimizes the level of effort associated with management of reservations. Three services that can help you standardize your instances are:

- [AWS Config](#) – Enables you to assess, audit, and evaluate the configurations of your AWS resources. AWS Config continuously monitors and records your AWS resource configurations and lets you automate the evaluation of recorded configurations against desired configurations.
- [Service Catalog](#) – Lets you create and manage catalogs of IT services that are approved for use on AWS. These IT services can include everything from virtual machine (VM) images, servers, software, and databases to complete multi-tier application architecture.
- [AWS Compute Optimizer](#) - Recommends optimal AWS compute resources for your workloads to reduce costs and improve performance by using Machine Learning algorithms to analyze historical utilization metrics. The Compute Optimizer focuses on the configuration and resource utilization of your workload to identify dozens of defining characteristics, such as whether a workload is CPU-intensive, exhibits a daily pattern, or accesses local storage frequently. The service processes these characteristics and identifies the hardware resource headroom required by the workload. It also infers how the workload would have performed on various hardware platforms (for example, Amazon EC2 instances types) and offers recommendations.

Reservation management techniques

You can manage reservations either by using a central IT operations or management team or by using a specific team or business unit. The following table summarizes the different reservation management techniques.

Table 5 – Comparison of different reservation management techniques

Central reservation management	Team/Business Unit reservation management
Maximizes reservation coverage by covering aggregate usage across a business	Increases likelihood of high reservation utilization (for example, using already-purchased reservations), because a single team should understand its capacity commitment of RIs
Simplifies overall reservation management especially when combining central management and Convertible Reserved Instances	Reduces interfacing or planning between the business unit and the central team

Central reservation management	Team/Business Unit reservation management
Reduces the requirement for an individual team to understand reservations	Streamlines decisions about purchases, purchase process, and reservation account location

Reserved Instance Marketplace

[Reserved Instance Marketplace](#) supports the sale of third-party and AWS customers' unused Standard Reserved Instances, which vary in term lengths and pricing options. For example, you might 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. For more information, see [Reserved Instance Marketplace](#).

AWS Cost Explorer

[AWS Cost Explorer](#) lets you visualize, understand, and manage your AWS costs and usage over time. You can analyze your cost and usage data at a high level (for example, total costs and usage across all accounts in your organization) or for highly specific requests (for example, m2.2xlarge costs within account Y that are tagged *project: secretProject*).

You can dive deeper into your reservations using the Reserved Instance utilization and coverage reports. Using these reports, you can set custom Reserved Instance utilization and coverage targets, and visualize progress toward your goals. From there, you can refine the underlying data using the available filtering dimensions (for example, account, instance type, scope, and more).

AWS Cost Explorer provides the following prebuilt reports:

- [EC2 RI Utilization %](#) offers relevant data to identify and act on opportunities to increase your Reserved Instance usage efficiency. It's calculated by dividing Reserved Instance hours used by the total Reserved Instance purchased hours.
- [EC2 RI Coverage %](#) shows how much of your overall instance usage is covered by Reserved Instances. This lets you make informed decisions about when to purchase or modify a Reserved

Instance to ensure maximum coverage. It's calculated by dividing Reserved Instance hours used by the total EC2 On-Demand and Reserved Instance hours.

Also, AWS Cost Explorer provides Reserved Instance purchase recommendations for zonal and size-flexible Reserved Instances to help payer accounts achieve greater cost efficiencies. For more information, see [AWS Cost Explorer](#).

AWS Cost and Usage Report

The [AWS Cost and Usage Report](#) contains the most comprehensive set of data about your AWS costs and usage, including additional information regarding AWS services, pricing, and reservations. By using the AWS Cost and Usage report, you can gain a wealth of reservation-related insights about the Amazon Resource Name (ARN) for a reservation, the number of reservations, the number of units per reservation, and more. It can help you do the following:

- Calculate savings** – Each hourly line item of usage contains the discounted rate that was charged, in addition to the public On-Demand Instance rate for that usage type at that time. You can quantify your savings by calculating the difference between the public On-Demand Instance rates and the rates you were charged.
- Track the allocation of Reserved Instance discounts** – Each line item of usage that receives a discount contains information about where the discount came from. This makes it easier to trace which instances are benefitting from specific reservations.

These reports update up to three times per day.

Reserved Instances on your cost and usage report

The Fee line item is added to your bill when you purchase an All Upfront or Partial Upfront Reserved Instance, as shown.

lineItem/ LineItemType	lineItem/ Product Code	lineItem/ UsageStartDate	lineItem/ Description	lineItem/ Unblended Cost	reservation/ ReservationARN
Fee	Amazon EC2	2016-01-01T00:00:00Z	Sign up charge for subscription: 363836886, planId: 1026576	68	arn:aws:ec2:us-east-1:171481847471:reserved-instances/1-00000-4371-with-1-171481847471

Figure 5 – Fee line item from AWS Cost and Usage Report

The **RI Fee** line item describes the recurring monthly charges that are associated with Partial Upfront and No Upfront Reserved Instances. The **RI Fee** is calculated by multiplying your discounted hourly rate by the number of hours in the month, as shown.

lineitem/ LineitemType	lineitem/ Product Code	lineitem/ UsageStartDate	lineitem/ UsageType	lineitem/ Description	lineitem/ Normalization Factor	lineitem/ Unblended Cost	reservation/ AvailabilityZone	reservation/ ReservationARN	reservation/ TotalReserved Units	reservation/ TotalReserved NormalizedUnits
Ri fee	Amazon EC2	2016-01-01T00:00:00Z	HeavyUsage: m4.large	USD 0.0309 hourly fee per Linux/Unix (Amazon VPC) m4.large instance	4	23	Amazwsc2us-east-1: Greserved-instances/ ----- -rding-----	744	2976	

Figure 6 – RI Fee line item from AWS Cost and Usage Report

The **Discounted Usage** line item describes the instance usage that received a matching Reserved Instance discount benefit. It's added to your bill when you have usage that matches one of your Reserved Instances, as shown.

lineitem/ LineitemType	lineitem/ Product Code	lineitem/ UsageStartDate	lineitem/ UsageType	lineitem/ Description	lineitem/ ResourceId	lineitem/ AvailabilityZone	lineitem/ UsageAmount	lineitem/ NormalizedUsageAmount	lineitem/ Unblended Rate	lineitem/ Unblended Cost	reservation/ ReservationARN
Discounted Usage	Amazon EC2	2016-01-01T00:00:00Z	BoxUsage: m4.large	Linux/Unix (Amazon VPC). m4.large reserved instance applied	-100000bc	us-east-1b	1	4	4	0	Amazwsc2us-east-1: Greserved-instances/ ----- -rding-----

Figure 7 – Discounted Usage line item from AWS Cost and Usage Report

AWS Trusted Advisor

[AWS Trusted Advisor](#) is an online resource to help you reduce cost, increase performance, and improve security by optimizing your AWS environment. AWS Trusted Advisor provides real-time guidance to help you provision your resources following AWS best practices. To help you maximize utilization of Reserved Instances, AWS Trusted Advisor checks your Amazon EC2 computing-consumption history and calculates an optimal number of Partial Upfront Reserved Instances. Recommendations are based on the previous calendar month's hour-by-hour usage aggregated across all consolidated billing accounts. Note that Trusted Advisor does not provide size-flexible Reserved Instance recommendations.

For more information about how the recommendation is calculated, see "Reserved Instance Optimization Check Questions" in the Trusted Advisor FAQs.

Conclusion

Effectively planned and managed, reservations can help you achieve significant discounts for AWS workloads that run on a predictable schedule. It's important to analyze your current AWS usage to select the right reservation attributes from the start and to devise a longer-term strategy for monitoring and managing your Reserved Instances. Using tools such as the AWS Compute Optimizer, AWS Cost and Usage report, and the Reserved Instance Utilization and Coverage reports in AWS Cost Explorer, you can examine your overall usage and discover opportunities for greater cost efficiencies.

Contributors

Contributors to this document include:

- Pritam Pal, Senior Specialist Solution Architect, EC2 Spot, Amazon Web Services

Document revisions

To be notified about updates to this whitepaper, subscribe to the RSS feed.

Change	Description	Date
<u>Updated bare metal instance types and normalization factors. Removed link to Scheduled Instances.</u>	Minor update.	March 29, 2021
<u>Updated Reserved Instances billing information, and normalization factors. Savings Plan section added.</u>	Whitepaper updated.	August 31, 2020
<u>Initial publication</u>	Whitepaper published.	March 1, 2018

Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided "as is" without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

© 2021 Amazon Web Services, Inc. or its affiliates. All rights reserved.