# Table of Contents

What Is ElastiCache for Memcached? ................................................................. 1
  Common ElastiCache Use Cases and How ElastiCache Can Help ................. 2
    In-Memory Data Store............................................................................. 2
    ElastiCache Customer Testimonials....................................................... 3
ElastiCache for Memcached Resources .......................................................... 4
ElastiCache for Memcached Components and Features ............................... 6
  Nodes ......................................................................................... 6
  Clusters .................................................................................... 7
AWS Regions and Availability Zones ........................................................... 8
Endpoints ............................................................................................ 8
Parameter Groups .................................................................................. 8
Security ......................................................................................... 8
Security Groups ................................................................................ 9
Subnet Groups ................................................................................... 9
Events ............................................................................................. 9

Tools for Managing Your Implementation .................................................... 10
  Using the AWS Management Console .................................................. 10
  Using the AWS CLI ......................................................................... 10
  Using the AWS SDK ...................................................................... 10
  Using the ElastiCache API ............................................................... 10
See also ............................................................................................ 10

Comparing Memcached and Redis ............................................................... 11
Getting Started with ElastiCache for Memcached ....................................... 15
Setting Up .......................................................................................... 15
  Create Your AWS Account ................................................................ 15
  Set Up Your Permissions (New ElastiCache Users Only) ..................... 15
Deploy a Memcached Cluster .................................................................... 16
  Determine Your Cluster's Requirements .......................................... 16
  Step 1: Launch a Memcached Cluster ............................................. 17
  Step 2: Authorize Access ................................................................ 19
  Step 3: Connect to a Cluster's Node .............................................. 20
  Step 4: Delete Your Cluster (Avoid Unnecessary Charges) ............... 21

Where Do I Go From Here? ...................................................................... 22

Tutorials and Videos ................................................................................ 23
  Videos ....................................................................................... 24
    Introductory Videos .................................................................. 24
    Advanced Videos .................................................................... 24
Caching Strategies and Best Practices ....................................................... 26
  Caching Strategies ..................................................................... 26
  Lazy Loading ............................................................................. 26
  Write-Through .......................................................................... 28
  Adding TTL .............................................................................. 29
  Related Topics ......................................................................... 30
Configuring Your ElastiCache Client for Efficient Load Balancing ............. 31
  Consistent Hashing Using Java .................................................... 31
  Consistent Hashing Using PHP ..................................................... 31
  Consistent Hashing Using .NET .................................................... 32
Managing Your ElastiCache for Memcached Implementation ..................... 33
  Engine Versions and Upgrading ....................................................... 33
    Supported Memcached Versions .............................................. 34
    Upgrading Engine Versions ..................................................... 36
  Choosing Regions and Availability Zones ....................................... 37
    Supported Regions & Endpoints ................................................ 38
    Locating Your Nodes .................................................................. 41

API Version 2015-02-02
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing Maintenance</td>
<td>41</td>
</tr>
<tr>
<td>Managing Nodes</td>
<td>42</td>
</tr>
<tr>
<td>Connecting to Nodes</td>
<td>43</td>
</tr>
<tr>
<td>Reserved Nodes</td>
<td>45</td>
</tr>
<tr>
<td>Supported Node Types</td>
<td>46</td>
</tr>
<tr>
<td>Replacing Nodes</td>
<td>50</td>
</tr>
<tr>
<td>Managing Your ElastiCache Clusters</td>
<td>51</td>
</tr>
<tr>
<td>Supported Memcached Versions</td>
<td>51</td>
</tr>
<tr>
<td>Other ElastiCache Cluster Operations</td>
<td>51</td>
</tr>
<tr>
<td>Creating a Cluster</td>
<td>53</td>
</tr>
<tr>
<td>Viewing a Cluster's Details</td>
<td>86</td>
</tr>
<tr>
<td>Modifying a Cluster</td>
<td>91</td>
</tr>
<tr>
<td>Rebooting a Cluster</td>
<td>93</td>
</tr>
<tr>
<td>Adding Nodes to a Cluster</td>
<td>94</td>
</tr>
<tr>
<td>Removing Nodes from a Cluster</td>
<td>99</td>
</tr>
<tr>
<td>Canceling Pending Add or Delete Node Operations</td>
<td>104</td>
</tr>
<tr>
<td>Deleting a Cluster</td>
<td>105</td>
</tr>
<tr>
<td>Accessing Your Cluster</td>
<td>107</td>
</tr>
<tr>
<td>Determine the Cluster's Platform</td>
<td>107</td>
</tr>
<tr>
<td>Grant Access to Your Cluster</td>
<td>109</td>
</tr>
<tr>
<td>Replication Across AWS Regions Using Global Datastore</td>
<td>115</td>
</tr>
<tr>
<td>Overview</td>
<td>115</td>
</tr>
<tr>
<td>Prerequisites and Limitations</td>
<td>116</td>
</tr>
<tr>
<td>Using Global Datastores (Console)</td>
<td>116</td>
</tr>
<tr>
<td>Using Global Datastores (CLI)</td>
<td>124</td>
</tr>
<tr>
<td>Finding Connection Endpoints</td>
<td>126</td>
</tr>
<tr>
<td>Finding Endpoints (Console)</td>
<td>128</td>
</tr>
<tr>
<td>Finding Endpoints (AWS CLI)</td>
<td>130</td>
</tr>
<tr>
<td>Finding Endpoints (ElastiCache API)</td>
<td>132</td>
</tr>
<tr>
<td>Scaling ElastiCache for Memcached Clusters</td>
<td>132</td>
</tr>
<tr>
<td>Scaling Memcached Horizontally</td>
<td>133</td>
</tr>
<tr>
<td>Scaling Memcached Vertically</td>
<td>133</td>
</tr>
<tr>
<td>Configuring Engine Parameters Using Parameter Groups</td>
<td>135</td>
</tr>
<tr>
<td>Parameter Management</td>
<td>136</td>
</tr>
<tr>
<td>Parameter Group Tiers</td>
<td>137</td>
</tr>
<tr>
<td>Creating a Parameter Group</td>
<td>137</td>
</tr>
<tr>
<td>Listing Parameter Groups by Name</td>
<td>141</td>
</tr>
<tr>
<td>Listing a Parameter Group's Values</td>
<td>145</td>
</tr>
<tr>
<td>Modifying a Parameter Group</td>
<td>146</td>
</tr>
<tr>
<td>Deleting a Parameter Group</td>
<td>149</td>
</tr>
<tr>
<td>Memcached Specific Parameters</td>
<td>151</td>
</tr>
<tr>
<td>Monitoring Usage, Events, and Costs</td>
<td>161</td>
</tr>
<tr>
<td>Monitoring Use</td>
<td>162</td>
</tr>
<tr>
<td>Host-Level Metrics</td>
<td>162</td>
</tr>
<tr>
<td>Metrics for Memcached</td>
<td>163</td>
</tr>
<tr>
<td>Which Metrics Should I Monitor?</td>
<td>166</td>
</tr>
<tr>
<td>Choosing Metric Statistics and Periods</td>
<td>167</td>
</tr>
<tr>
<td>Monitoring CloudWatch Cluster and Node Metrics</td>
<td>167</td>
</tr>
<tr>
<td>Monitoring Events</td>
<td>170</td>
</tr>
<tr>
<td>Managing ElastiCache Amazon SNS Notifications</td>
<td>170</td>
</tr>
<tr>
<td>Viewing ElastiCache Events</td>
<td>173</td>
</tr>
<tr>
<td>Event Notifications and Amazon SNS</td>
<td>175</td>
</tr>
<tr>
<td>Monitoring Costs with Tags</td>
<td>179</td>
</tr>
<tr>
<td>Managing Tags Using the Console</td>
<td>180</td>
</tr>
<tr>
<td>Managing Tags Using the AWS CLI</td>
<td>181</td>
</tr>
<tr>
<td>Managing Tags Using the ElastiCache API</td>
<td>183</td>
</tr>
<tr>
<td>Managing Costs with Reserved Nodes</td>
<td>186</td>
</tr>
</tbody>
</table>

API Version 2015-02-02
What Is Amazon ElastiCache for Memcached?

Welcome to the Amazon ElastiCache for Memcached User Guide. Amazon ElastiCache is a web service that makes it easy to set up, manage, and scale a distributed in-memory data store or cache environment in the cloud. It provides a high-performance, scalable, and cost-effective caching solution. At the same time, it helps remove the complexity associated with deploying and managing a distributed cache environment.

**Note**
Amazon ElastiCache works with both the Redis and Memcached engines. Use the guide for the engine that you're interested in. If you're unsure which engine you want to use, see Comparing Memcached and Redis (p. 11) in this guide.

Existing applications that use Memcached can use ElastiCache with almost no modification. Your applications simply need information about the host names and port numbers of the ElastiCache nodes that you have deployed. The ElastiCache Auto Discovery feature for Memcached lets your applications identify all of the nodes in a cache cluster and connect to them. This means that you don't have to maintain a list of available host names and port numbers. In this way, your applications are effectively insulated from changes to node membership in a cluster.

ElastiCache for Memcached has multiple features to enhance reliability for critical production deployments:

- Automatic detection and recovery from cache node failures.
- Automatic discovery of nodes within a cluster enabled for automatic discovery, so that no changes need to be made to your application when you add or remove nodes.
- Flexible Availability Zone placement of nodes and clusters.
- Integration with other AWS services such as Amazon EC2, Amazon CloudWatch, AWS CloudTrail, and Amazon SNS to provide a secure, high-performance, managed in-memory caching solution.

**Topics**
- Common ElastiCache Use Cases and How ElastiCache Can Help (p. 2)
- ElastiCache for Memcached Resources (p. 4)
- ElastiCache for Memcached Components and Features (p. 6)
- Tools for Managing Your Implementation (p. 10)
Common ElastiCache Use Cases and How ElastiCache Can Help

Whether serving the latest news or a product catalog, or selling tickets to an event, speed is the name of the game. The success of your website and business is greatly affected by the speed at which you deliver content.

In "For Impatient Web Users, an Eye Blink Is Just Too Long to Wait," the New York Times noted that users can register a 250-millisecond (1/4 second) difference between competing sites. Users tend to opt out of the slower site in favor of the faster site. Tests done at Amazon, cited in How Webpage Load Time Is Related to Visitor Loss, revealed that for every 100-ms (1/10 second) increase in load time, sales decrease 1 percent.

If someone wants data, you can deliver that data much faster if it's cached. That's true whether it's for a webpage or a report that drives business decisions. Can your business afford to not cache your webpages so as to deliver them with the shortest latency possible?

It might seem intuitively obvious that you want to cache your most heavily requested items. But why not cache your less frequently requested items? Even the most optimized database query or remote API call is noticeably slower than retrieving a flat key from an in-memory cache. Noticeably slower tends to send customers elsewhere.

The following examples illustrate some of the ways using ElastiCache can improve overall performance of your application.

**In-Memory Data Store**

The primary purpose of an in-memory key-value store is to provide ultrafast (submillisecond latency) and inexpensive access to copies of data. Most data stores have areas of data that are frequently accessed but seldom updated. Additionally, querying a database is always slower and more expensive than locating a key in a key-value pair cache. Some database queries are especially expensive to perform. An example is queries that involve joins across multiple tables or queries with intensive calculations. By caching such query results, you pay the price of the query only once. Then you can quickly retrieve the data multiple times without having to re-execute the query.

The following image shows ElastiCache caching.
What Should I Cache?

When deciding what data to cache, consider these factors:

**Speed and expense** – It’s always slower and more expensive to get data from a database than from a cache. Some database queries are inherently slower and more expensive than others. For example, queries that perform joins on multiple tables are much slower and more expensive than simple, single table queries. If the interesting data requires a slow and expensive query to get, it’s a candidate for caching. If getting the data requires a relatively quick and simple query, it might still be a candidate for caching, depending on other factors.

**Data and access pattern** – Determining what to cache also involves understanding the data itself and its access patterns. For example, it doesn’t make sense to cache data that changes quickly or is seldom accessed. For caching to provide a real benefit, the data should be relatively static and frequently accessed. An example is a personal profile on a social media site. On the other hand, you don’t want to cache data if caching it provides no speed or cost advantage. For example, it doesn’t make sense to cache webpages that return search results because the queries and results are usually unique.

**Staleness** – By definition, cached data is stale data. Even if in certain circumstances it isn’t stale, it should always be considered and treated as stale. To tell whether your data is a candidate for caching, determine your application’s tolerance for stale data.

Your application might be able to tolerate stale data in one context, but not another. For example, suppose that your site serves a publicly traded stock price. Your customers might accept some staleness with a disclaimer that prices might be \( n \) minutes delayed. But if you serve that stock price to a broker making a sale or purchase, you want real-time data.

Consider caching your data if the following is true:

- Your data is slow or expensive to get when compared to cache retrieval.
- Users access your data often.
- Your data stays relatively the same, or if it changes quickly staleness is not a large issue.

For more information, see the following:

- **Caching Strategies** in the *ElastiCache for Memcached User Guide*

ElastiCache Customer Testimonials

To learn about how businesses like Airbnb, PBS, Esri, and others use Amazon ElastiCache to grow their businesses with improved customer experience, see *How Others Use Amazon ElastiCache*.

You can also watch the **ElastiCache Videos (p. 24)** for additional ElastiCache customer use cases.
ElastiCache for Memcached Resources

We recommend that you begin by reading the following sections, and refer to them as you need them:

- **Service Highlights and Pricing** – The product detail page provides a general product overview of ElastiCache, service highlights, and pricing.

- **ElastiCache Videos** – The ElastiCache Videos (p. 24) section has videos that introduce you to Amazon ElastiCache for Memcached, cover common use cases, and demo how to use ElastiCache for Memcached to reduce latency and improve throughput of your applications.

- **Getting Started** – The Getting Started with Amazon ElastiCache for Memcached (p. 15) section includes an example that walks you through the process of creating a cache cluster, authorizing access to the cache cluster, connecting to a cache node, and deleting the cache cluster.

- **Performance at Scale** – The Performance at Scale with Amazon ElastiCache white paper addresses caching strategies that enable your application to perform well at scale.

After you complete the preceding sections, read these sections:

- **Choosing Your Memcached Node Size** (p. 80)
  You want your nodes to be large enough to accommodate all the data you want to cache. At the same time, you don’t want to pay for more cache than you need. You can use this topic to help select the best node size.

- **Caching Strategies and Best Practices** (p. 26)
  Identify and address issues that can impact the efficiency of your cluster.

If you want to use the AWS Command Line Interface (AWS CLI), you can use these documents to help you get started:

- **AWS Command Line Interface Documentation**
  This section provides information on downloading the AWS CLI, getting the AWS CLI working on your system, and providing your AWS credentials.

- **AWS CLI Documentation for ElastiCache**
  This separate document covers all of the AWS CLI for ElastiCache commands, including syntax and examples.

You can write application programs to use the ElastiCache API with a variety of popular programming languages. Here are some resources:

- **Tools for Amazon Web Services**
  Amazon Web Services provides a number of software development kits (SDKs) with support for ElastiCache for Memcached. You can code for ElastiCache using Java, .NET, PHP, Ruby, and other languages. These SDKs can greatly simplify your application development by formatting your requests to ElastiCache, parsing responses, and providing retry logic and error handling.

- **Using the ElastiCache API** (p. 262)
  If you don’t want to use the AWS SDKs, you can interact with ElastiCache directly using the Query API. You can find troubleshooting tips and information on creating and authenticating requests and handling responses in this section.

- **Amazon ElastiCache API Reference**
This separate document covers all of the ElastiCache API operations, including syntax and examples.
ElastiCache for Memcached Components and Features

Following, you can find an overview of the major components of an Amazon ElastiCache for Memcached deployment.

Topics

- ElastiCache Nodes (p. 6)
- ElastiCache for Memcached Clusters (p. 7)
- AWS Regions and Availability Zones (p. 8)
- ElastiCache for Memcached Endpoints (p. 8)
- ElastiCache Parameter Groups (p. 8)
- ElastiCache Security (p. 8)
- ElastiCache Security Groups (p. 9)
- ElastiCache Subnet Groups (p. 9)
- ElastiCache for Memcached Events (p. 9)

ElastiCache Nodes

A *node* is the smallest building block of an ElastiCache deployment. A node can exist in isolation from or in some relationship to other nodes.

A node is a fixed-size chunk of secure, network-attached RAM. Each node runs an instance of Memcached. If necessary, you can scale the nodes in a cluster up or down to a different instance type. For more information, see Scaling ElastiCache for Memcached Clusters (p. 132).

Every node within a cluster is the same instance type and runs the same cache engine. Each cache node has its own Domain Name Service (DNS) name and port. Multiple types of cache nodes are supported, each with varying amounts of associated memory. For a list of supported node instance types, see Supported Node Types (p. 46).

You can purchase nodes on a pay-as-you-go basis, where you only pay for your use of a node. Or you can purchase reserved nodes at a significantly reduced hourly rate. If your usage rate is high, purchasing reserved nodes can save you money. Suppose that your cluster is almost always in use, and you occasionally add nodes to handle use spikes. In this case, you can purchase a number of reserved nodes to run most of the time and purchase pay-as-you-go nodes for the times you occasionally need to add nodes. For more information on reserved nodes, see ElastiCache Reserved Nodes (p. 45).

The Memcached engine supports Auto Discovery. *Auto Discovery* is the ability for client programs to automatically identify all of the nodes in a cache cluster, and to initiate and maintain connections to all of these nodes. With Auto Discovery, your application doesn't need to manually connect to individual nodes. Instead, your application connects to a configuration endpoint. The configuration endpoint DNS entry contains the CNAME entries for each of the cache node endpoints. Thus, by connecting to the configuration endpoint, your application immediately has information about all of the nodes in the cluster and can connect to all of them. You don't need to hard-code the individual cache node endpoints in your application. For more information on Auto Discovery, see Automatically Identify Nodes in your Memcached Cluster (p. 55).

For more information on nodes, see Managing Nodes (p. 42).
ElastiCache for Memcached Clusters

A Memcached cluster is a logical grouping of one or more ElastiCache Nodes (p. 6). Data is partitioned across the nodes in a Memcached cluster.

Many ElastiCache operations are targeted at clusters:

- Creating a cluster
- Modifying a cluster
- Deleting a cluster
- Viewing the elements in a cluster
- Adding or removing cost allocation tags to and from a cluster

For more detailed information, see the following related topics:

- Managing Your ElastiCache Clusters (p. 51) and Managing Nodes (p. 42)
  Information about clusters, nodes, and related operations.
- AWS Service Limits: Amazon ElastiCache
  Information about ElastiCache limits, such as the maximum number of nodes or clusters.

If you need to exceed these limits, make your request using the Amazon ElastiCache Cache Node request form.

- Mitigating Failures (p. 251)
  Information about improving the fault tolerance of your clusters.

Typical Cluster Configurations

Memcached supports up to 100 nodes per customer for each AWS Region with each cluster having 1–20 nodes. You partition your data across the nodes in a Memcached cluster.

When you run the Memcached engine, clusters can be made up of 1–20 nodes. You partition your database across the nodes. Your application reads and writes to each node's endpoint. For more information, see Automatically Identify Nodes in your Memcached Cluster (p. 55).

For improved fault tolerance, locate your Memcached nodes in various Availability Zones (AZs) within the cluster's AWS Region. That way, a failure in one AZ has minimal impact upon your entire cluster and application. For more information, see Mitigating Failures (p. 251).

As demand upon your Memcached cluster changes, you can scale out or in by adding or removing nodes, which repartitions your data across the new number of nodes. When you partition your data, we recommend using consistent hashing. For more information about consistent hashing, see Configuring Your ElastiCache Client for Efficient Load Balancing (p. 31). In the following diagram, you can see examples of single node and multiple node Memcached clusters.
AWS Regions and Availability Zones

Amazon ElastiCache for Memcached is available in multiple AWS Regions around the world. Thus, you can launch ElastiCache clusters in the locations that meet your business requirements. For example, you can launch in the AWS Region closest to your customers or to meet certain legal requirements.

By default, the AWS SDKs, AWS CLI, ElastiCache API, and ElastiCache console reference the US-West (Oregon) region. As ElastiCache expands availability to new AWS Regions, new endpoints for these AWS Regions are also available to use in your HTTP requests, the AWS SDKs, AWS CLI, and ElastiCache console.

Each AWS Region is designed to be completely isolated from the other AWS Regions. Within each are multiple Availability Zones. By launching your nodes in different Availability Zones, you can achieve the greatest possible fault tolerance. For more information about AWS Regions and Availability Zones, see Choosing Regions and Availability Zones (p. 37).

ElastiCache for Memcached Endpoints

An endpoint is the unique address your application uses to connect to an ElastiCache node or cluster.

Each node in a Memcached cluster has its own endpoint. The cluster also has an endpoint called the configuration endpoint. If you enable Auto Discovery and connect to the configuration endpoint, your application automatically knows each node endpoint, even after adding or removing nodes from the cluster. For more information, see Automatically Identify Nodes in your Memcached Cluster (p. 55).

ElastiCache Parameter Groups

Cache parameter groups are an easy way to manage runtime settings for supported engine software. Parameters are used to control memory usage, eviction policies, item sizes, and more. An ElastiCache parameter group is a named collection of engine-specific parameters that you can apply to a cluster. By doing this, you make sure that all of the nodes in that cluster are configured in exactly the same way.

For a list of supported parameters, their default values, and which ones can be modified, see DescribeEngineDefaultParameters (describe-engine-default-parameters).

For more detailed information on ElastiCache parameter groups, see Configuring Engine Parameters Using Parameter Groups (p. 135).

ElastiCache Security

For enhanced security, ElastiCache node access is restricted to applications running on whitelisted Amazon EC2 instances. You can control the Amazon EC2 instances that can access your cluster by using subnet groups or security groups.
By default, all new ElastiCache clusters are launched in an Amazon Virtual Private Cloud (Amazon VPC) environment. You can use subnet groups to grant cluster access from Amazon EC2 instances running on specific subnets. If you choose to run your cluster outside of Amazon VPC, you can create security groups to authorize Amazon EC2 instances running within specific Amazon EC2 security groups.

### ElastiCache Security Groups

**Note**

ElastiCache security groups are only applicable to clusters that are not running in an Amazon Virtual Private Cloud (Amazon VPC) environment. If you are running your ElastiCache nodes in an Amazon VPC, you control access to your cache clusters with Amazon VPC security groups, which are different from ElastiCache security groups.

For more information on using ElastiCache in an Amazon VPC, see Amazon VPCs and ElastiCache Security (p. 198).

ElastiCache allows you to control access to your clusters using security groups. A security group acts like a firewall, controlling network access to your cluster. By default, network access to your clusters is turned off. If you want your applications to access your cluster, you must explicitly enable access from hosts in specific Amazon EC2 security groups. After ingress rules are configured, the same rules apply to all clusters associated with that security group.

To allow network access to your cluster, first create a security group. Then use the AuthorizeCacheSecurityGroupIngress API action or the authorize-cache-security-group-ingress AWS CLI command to authorize the desired Amazon EC2 security group. Doing this in turn specifies the Amazon EC2 instances allowed. You can associate the security group with your cluster at the time of creation, or by using the ElastiCache management console or the ModifyCacheCluster or (modify-cache-cluster) AWS CLI for ElastiCache command.

**Important**

IP-range based access control is currently not enabled for clusters. All clients to a cluster must be within the Amazon EC2 network, and authorized via security groups as described previously.

For more information about security groups, see Security Groups: EC2-Classic (p. 222).

### ElastiCache Subnet Groups

A subnet group is a collection of subnets (typically private) that you can designate for your clusters running in an Amazon Virtual Private Cloud (Amazon VPC) environment.

If you create a cluster in an Amazon VPC, then you must specify a cache subnet group. ElastiCache uses that cache subnet group to choose a subnet and IP addresses within that subnet to associate with your cache nodes.

For more information about cache subnet group usage in an Amazon VPC environment, see Amazon VPCs and ElastiCache Security (p. 198), Step 2: Authorize Access (p. 19), and Subnets and Subnet Groups (p. 214).

### ElastiCache for Memcached Events

When significant events happen on a cache cluster, ElastiCache sends notification to a specific Amazon SNS topic. Significant events can include such things as a failure to add a node, success in adding a node, the modification of a security group, and others. By monitoring for key events, you can know the current state of your clusters and, depending upon the event, take corrective action.

For more information on ElastiCache events, see Monitoring ElastiCache Events (p. 170).
Tools for Managing Your Implementation

Once you have granted your Amazon EC2 instance access to your ElastiCache cluster, you have four means by which you can manage your ElastiCache cluster: the AWS Management Console, the AWS CLI for ElastiCache, the AWS SDK for ElastiCache, and the ElastiCache API.

Using the AWS Management Console

The AWS Management Console is the easiest way to manage Amazon ElastiCache for Memcached. The console lets you create cache clusters, add and remove cache nodes, and perform other administrative tasks without having to write any code. The console also provides cache node performance graphs from CloudWatch, showing cache engine activity, memory and CPU utilization, as well as other metrics. For more information, see specific topics in this User Guide.

Using the AWS CLI

You can also use the AWS Command Line Interface (AWS CLI) for ElastiCache. The AWS CLI makes it easy to perform one-at-a-time operations, such as starting or stopping your cache cluster. You can also invoke AWS CLI for ElastiCache commands from a scripting language of your choice, letting you automate repeating tasks. For more information about the AWS CLI, see the User Guide and the AWS CLI Command Reference.

Using the AWS SDK

If you want to access ElastiCache from an application, you can use one of the AWS software development kits (SDKs). The SDKs wrap the ElastiCache API calls, and insulate your application from the low-level details of the ElastiCache API. You provide your credentials, and the SDK libraries take care of authentication and request signing. For more information about using the AWS SDKs, see Tools for Amazon Web Services.

Using the ElastiCache API

You can also write application code directly against the ElastiCache web service API. When using the API, you must write the necessary code to construct and authenticate your HTTP requests, parse the results from ElastiCache, and handle any errors. For more information about the API, see Using the ElastiCache API (p. 262).

See also

For more detailed information on managing your Amazon ElastiCache for Memcached deployment, see the following:

- Managing Your ElastiCache for Memcached Implementation (p. 33)
- Internetwork Traffic Privacy (p. 198)
- Monitoring Usage, Events, and Costs (p. 161)
Comparing Memcached and Redis

Amazon ElastiCache supports the Memcached and Redis cache engines. Each engine provides some advantages. Use the information in this topic to help you choose the engine and version that best meets your requirements.

**Important**
After you create a cache cluster or replication group, you can upgrade to a newer engine version, but you cannot downgrade to an older engine version. If you want to use an older engine version, you must delete the existing cache cluster or replication group and create it again with the earlier engine version.

On the surface, the engines look similar. Each of them is an in-memory key-value store. However, in practice there are significant differences.

**Choose Memcached if the following apply for you:**
- You need the simplest model possible.
- You need to run large nodes with multiple cores or threads.
- You need the ability to scale out and in, adding and removing nodes as demand on your system increases and decreases.
- You need to cache objects, such as a database.

**Choose Redis with a version of ElastiCache for Redis if the following apply for you:**
- **ElastiCache for Redis version 5.0.0 (Enhanced)**
  You want to use Redis streams, a log data structure that allows producers to append new items in real time and also allows consumers to consume messages either in a blocking or non-blocking fashion.
  
  For more information, see Redis Version 5.0.0 (Enhanced).
- **ElastiCache for Redis version 4.0.10 (Enhanced)**
  Supports both encryption and dynamically adding or removing shards from your Redis (cluster mode enabled) cluster.
  
  For more information, see Redis Version 4.0.10 (Enhanced).
- **ElastiCache for Redis version 3.2.10 (Enhanced)**
  Supports the ability to dynamically add or remove shards from your Redis (cluster mode enabled) cluster.
  
  **Important**
  Currently ElastiCache for Redis 3.2.10 doesn’t support encryption.
  
  For more information, see the following:
  - Redis Version 3.2.10 (Enhanced)
  - Online resharding best practices for Redis, For more information, see the following:
    - Best Practices: Online Resharding
    - Online Resharding and Shard Rebalancing for Redis (Cluster Mode Enabled)
  - For more information on scaling Redis clusters, see Scaling.
- **ElastiCache for Redis version 3.2.6 (Enhanced)**
If you need the functionality of earlier Redis versions plus the following features, choose ElastiCache for Redis 3.2.6:

- **In-transit encryption.** For more information, see Amazon ElastiCache for Redis In-Transit Encryption.
- **At-rest encryption.** For more information, see Amazon ElastiCache for Redis At-Rest Encryption.
- **HIPAA eligibility certification.** For more information, see HIPAA Eligibility for Amazon ElastiCache for Redis.

**ElastiCache for Redis (Cluster mode enabled) version 3.2.4**

If you need the functionality of Redis 2.8.x plus the following features, choose Redis 3.2.4 (clustered mode):

- You need to partition your data across two to 90 node groups (clustered mode only).
- You need geospatial indexing (clustered mode or non-clustered mode).
- You don’t need to support multiple databases.

**Important**

Redis (cluster mode enabled) has the following limitations:

- No scale-up to larger node types
- No changing the number of replicas in a node group (partition)

**ElastiCache for Redis (non-clustered mode) 2.8x and 3.2.4 (Enhanced)**

If the following apply for you, choose Redis 2.8.x or Redis 3.2.4 (non-clustered mode):

- You need complex data types, such as strings, hashes, lists, sets, sorted sets, and bitmaps.
- You need to sort or rank in-memory datasets.
- You need persistence of your key store.
- You need to replicate your data from the primary to one or more read replicas for read intensive applications.
- You need automatic failover if your primary node fails.
- You need publish and subscribe (pub/sub) capabilities—to inform clients about events on the server.
- You need backup and restore capabilities.
- You need to support multiple databases.
## Comparison summary of Memcached, Redis (cluster mode disabled), and Redis (cluster mode enabled)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Memcached</th>
<th>Redis (cluster mode disabled)</th>
<th>Redis (cluster mode enabled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engine versions</td>
<td>1.4.x</td>
<td>2.8.x and later</td>
<td>3.2.x and later</td>
</tr>
<tr>
<td>Data types</td>
<td>Simple</td>
<td>2.8.x - Complex *</td>
<td>3.2.x and later - Complex</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complex</td>
<td></td>
</tr>
<tr>
<td>Data partitioning</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Cluster is modifiable</td>
<td>Yes</td>
<td>Yes</td>
<td>3.2.10 and later - Limited</td>
</tr>
<tr>
<td>Online resharding</td>
<td>No</td>
<td>No</td>
<td>3.2.10 and later</td>
</tr>
<tr>
<td>Encryption</td>
<td>No</td>
<td>3.2.6, 4.0.10 and later</td>
<td>3.2.6, 4.0.10 and later</td>
</tr>
<tr>
<td>Compliance certifications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FedRAMP</td>
<td>No</td>
<td>3.2.6, 4.0.10 and later</td>
<td>3.2.6, 4.0.10 and later</td>
</tr>
<tr>
<td>HIPAA</td>
<td>No</td>
<td>3.2.6, 4.0.10 and later</td>
<td>3.2.6, 4.0.10 and later</td>
</tr>
<tr>
<td>PCI DSS</td>
<td>No</td>
<td>3.2.6, 4.0.10 and later</td>
<td>3.2.6, 4.0.10 and later</td>
</tr>
<tr>
<td>Multi-threaded</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Node type upgrade</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Engine upgrading</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>High availability (replication)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Automatic failover</td>
<td>No</td>
<td>Optional</td>
<td>Required</td>
</tr>
<tr>
<td>Pub/Sub capabilities</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sorted sets</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Backup and restore</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Geospatial indexing</td>
<td>No</td>
<td>2.8.x - No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.2.x and later - Yes</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

- string, objects (like databases)
- * string, sets, sorted sets, lists, hashes, bitmaps, hyperloglog
- string, sets, sorted sets, lists, hashes, bitmaps, hyperloglog, geospatial indexes
After you choose the engine for your cluster, we recommend that you use the most recent version of that engine. For more information, see Supported ElastiCache for Memcached Versions or Supported ElastiCache for Redis Versions.
Getting Started with Amazon ElastiCache for Memcached

The topics in this section walk you through the process of creating, granting access to, connecting to, and finally deleting a Memcached cluster using the ElastiCache console.

Topics
- Setting Up (p. 15)
- Deploy a Memcached Cluster (p. 16)
- Where Do I Go From Here? (p. 22)

Setting Up

Following, you can find topics that describe the one-time actions you must take to start using ElastiCache.

Topics
- Create Your AWS Account (p. 15)
- Set Up Your Permissions (New ElastiCache Users Only) (p. 15)

Create Your AWS Account

To use Amazon ElastiCache, you must have an active AWS account and permissions to access ElastiCache and other AWS resources.

If you don't already have an AWS account, create one now. AWS accounts are free. You are not charged for signing up for an AWS service, only for using AWS services.

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.

Set Up Your Permissions (New ElastiCache Users Only)

Amazon ElastiCache creates and uses service-linked roles to provision resources and access other AWS resources and services on your behalf. For ElastiCache to create a service-linked role for you, use the AWS-managed policy named AmazonElastiCacheFullAccess. This role comes preprovisioned with permission that the service requires to create a service-linked role on your behalf.
You might decide not to use the default policy and instead to use a custom-managed policy. In this case, make sure that you have either permissions to call `iam:createServiceLinkedRole` or that you have created the ElastiCache service-linked role.

For more information, see the following:

- Creating a New Policy (IAM)
- AWS-Managed (Predefined) Policies for Amazon ElastiCache (p. 238)
- Using Service-Linked Roles for Amazon ElastiCache (p. 241)

**Deploy a Memcached Cluster**

Now that you have your AWS account and permissions set up, you can try out ElastiCache for Memcached by deploying a cluster. The following sections show you how to do this.

**Topics**

- Determine Your Cluster's Requirements (p. 16)
- Step 1: Launch a Memcached Cluster (p. 17)
- Step 2: Authorize Access (p. 19)
- Step 3: Connect to a Cluster's Node (p. 20)
- Step 4: Delete Your Cluster (Avoid Unnecessary Charges) (p. 21)

**Determine Your Cluster's Requirements**

Before you create a Memcached cluster, you should always determine the requirements for the cluster so that when you create the cluster it will meet your business needs and not need to be redone. Because in this exercise we will largely accept default values for the cluster, we will dispense with determining requirements. For more information, see Determine Your Requirements (p. 53).
Step 1: Launch a Memcached Cluster

The cluster you're about to launch will be live, and not running in a sandbox. You incur the standard ElastiCache usage fees for the instance until you delete it. The total charges are minimal (typically less than a dollar) if you complete the exercise described here in one sitting and delete your cluster when you are finished. For more information about ElastiCache usage rates, see https://aws.amazon.com/elasticache/.

Important
Your cluster is launched in an Amazon VPC. Before you start creating your cluster, you need to create a subnet group. For more information, see Creating a Subnet Group (p. 215).

To create an ElastiCache for Memcached cluster

2. Choose Get Started Now.
   
   If you already have an available cluster, choose Launch Cluster.
3. From the list in the upper-right corner, choose the AWS Region you want to launch this cluster in.
4. For Cluster engine, choose Memcached.
5. Complete the Memcached settings section as follows:
   
   a. In Name, type a name for your cluster.
      
      Cluster naming constraints are as follows:
      
      • Must contain 1–40 alphanumeric characters or hyphens.
      • Must begin with a letter.
      • Can’t contain two consecutive hyphens.
      • Can’t end with a hyphen.
   
   b. From the Engine version compatibility list, choose the Memcached engine version you want to run on this cluster. Unless you have a specific reason to run an older version, we recommend that you choose the latest version.
   
   c. In Port, accept the default port, 11211. If you have a reason to use a different port, type in the port number.
   
   d. From Parameter group, choose the parameter group you want to use with this cluster, or choose “Create new” to create a new parameter group to use with this cluster. For this exercise, accept the default parameter group.
      
      For more information, see Creating a Parameter Group (p. 137).
   
   e. For Node type, choose the node type that you want to use for this cluster. For this exercise, you can accept the default node type or select another node type.
      
      For more information, see Choosing Your Memcached Node Size (p. 80).
      
      To select another node type:
      
      i. Choose the down-arrow to the right of the default node type.
      
      ii. Choose the Instance family you want for the nodes in this cluster. Since this is just an exercise, to save costs, choose t2.
      
      iii. From the available node types, choose the box to the left of the node type you want for this cluster. Since this is just an exercise, to save costs, choose cache.t2.small.
   
   iv. Choose Save.
f. From the **Number of nodes** list, choose the number of nodes (partitions) you want provisioned for this cluster.

6. Choose **Advanced Memcached settings** and complete the section as follows:

   a. From the **Subnet group** list, choose the subnet you want to apply to this cluster. For this exercise, accept the default subnet group.

      For more information, see **Subnets and Subnet Groups** (p. 214).

   b. For **Availability zone(s)**, you have two options.

      • **No preference** – ElastiCache chooses each node's Availability Zone for you.

      • **Specify availability zones** – You specify the Availability Zone for each node.

      For this exercise, choose **Specify availability zones** and then choose an Availability Zone for each node.

      For more information, see **Choosing Regions and Availability Zones** (p. 37).

   c. From the **Security groups** list, choose the security groups that you want to use for this cluster. For this exercise, accept the default security group.

      For more information, see **Amazon VPCs and ElastiCache Security** (p. 198).

   d. The **Maintenance window** is the time, generally an hour, each week where ElastiCache schedules system maintenance on your cluster. You have two options.

      • **No preference**—ElastiCache chooses the day of the week and time of day for the cluster's maintenance window.

      • **Specify maintenance window**—You choose the day of the week, start time, and duration for the cluster's maintenance window.

      After the cluster is created, you can modify it to specify a difference maintenance window. For more information, see **Managing Maintenance** (p. 41).

   e. For **Notifications**, leave it as **Disable notifications**.

7. Choose **Create** to launch your cluster, or **Cancel** to cancel the operation.
Step 2: Authorize Access

This section assumes that you are familiar with launching and connecting to Amazon EC2 instances. For more information, see the *Amazon EC2 Getting Started Guide*.

All ElastiCache clusters are designed to be accessed from an Amazon EC2 instance. The most common scenario is to access an ElastiCache cluster from an Amazon EC2 instance in the same Amazon Virtual Private Cloud (Amazon VPC). This is the scenario covered in this topic. For information on accessing your ElastiCache cluster from a different Amazon VPC, a different region, or even your corporate network, see the following:

- *Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC* (p. 204)
- *Accessing ElastiCache Resources from Outside AWS* (p. 111)

By default, network access to your cluster is limited to the user account that was used to launch it. Before you can connect to a cluster from an EC2 instance, you must authorize the EC2 instance to access the cluster. The steps required depend upon whether you launched your cluster into EC2-VPC or EC2-Classic.

For the steps to authorize access to your cluster, see *Accessing Your Cluster* (p. 107).
Step 3: Connect to a Cluster's Node

Before you continue, be sure you have completed Step 2: Authorize Access (p. 19).

This section assumes that you've created an Amazon EC2 instance and can connect to it. For instructions on how to do this, see the Amazon EC2 Getting Started Guide.

An Amazon EC2 instance can connect to a cluster node only if you have authorized it to do so. For more information, see Step 2: Authorize Access (p. 19).

Step 3.1: Find Your Node Endpoints

When your cluster is in the available state and you've authorized access to it (Step 2: Authorize Access (p. 19)), you can log in to an Amazon EC2 instance and connect to the cluster. To do so, you must first determine the endpoint.

To find your endpoints, see the relevant topic from the following list. When you find the endpoint you need, copy it to your clipboard for use in Step 3.2.

- Finding Connection Endpoints (p. 126)
- Finding a Cluster's Endpoints (Console) (p. 128)—You need the cluster's Configuration endpoint.
- Finding Endpoints (AWS CLI) (p. 130)
- Finding Endpoints (ElastiCache API) (p. 132)

Step 3.2: Connect to a Memcached Cluster

Once your cluster is in the available state and you've authorized access to it (Step 2: Authorize Access (p. 19), you can log in to an Amazon EC2 instance and connect to the cluster.

For instructions on connecting to a Memcached cluster, see Connecting to Nodes (p. 43).
Step 4: Delete Your Cluster (Avoid Unnecessary Charges)

Important
It is almost always a good idea to delete clusters that you are not actively using. Until a cluster's status is deleted, you continue to incur charges for it.

Before you continue, be sure you have completed at least as far as Step 1: Launch a Memcached Cluster (p. 17).

To delete a Memcached cluster

1. Sign in to the AWS Management Console and open the Amazon ElastiCache console at https://console.aws.amazon.com/elasticache/.
2. To see a list of all your clusters running Memcached, in the navigation pane, choose Memcached.
3. To select the cluster to delete, select the cluster's name from the list of clusters.
   Tip
   You can only delete one cluster at a time from the ElastiCache console. Selecting multiple clusters disables the delete operation. To delete multiple clusters, repeat this process for each cluster.
4. For Actions, choose Delete.
5. In the Delete Cluster confirmation screen, choose Delete to delete the cluster, or Cancel to keep the cluster.
   If you choose Delete, the status of the cluster changes to deleting.

As soon as your cluster is no longer listed in the list of clusters, you stop incurring charges for it.

Now you have successfully launched, authorized access to, connected to, viewed, and deleted your first ElastiCache for Memcached cluster.
Where Do I Go From Here?

Now that you have tried the Getting Started exercise, you can explore the following sections to learn more about ElastiCache and available tools:

- Getting Started with AWS
- Tools for Amazon Web Services
- AWS Command Line Interface
- Amazon ElastiCache API Reference

After you complete the Getting Started exercise, you can read these sections to learn more about ElastiCache administration:

- Choosing Your Memcached Node Size (p. 80)
  
  You want your cache to be large enough to accommodate all the data you want to cache. At the same time, you don’t want to pay for more cache than you need. Use this topic to help you choose the best node size.

- Caching Strategies and Best Practices (p. 26)
  
  Identify and address issues that can affect the efficiency of your cluster.
ElastiCache Tutorials and Videos

The following tutorials address tasks of interest to the Amazon ElastiCache user.

- ElastiCache Videos (p. 24)
- Tutorial: Configuring a Lambda Function to Access Amazon ElastiCache in an Amazon VPC
ElastiCache Videos

Following, you can find videos to help you learn basic and advanced Amazon ElastiCache concepts. For information about AWS Training, see AWS Training & Certification.

Topics
- Introductory Videos (p. 24)
- Advanced Videos (p. 24)

Introductory Videos

The following videos introduce you to Amazon ElastiCache.

Topics
- DAT204—Building Scalable Applications on AWS NoSQL Services (re:Invent 2015) (p. 24)
- DAT207—Accelerating Application Performance with Amazon ElastiCache (AWS re:Invent 2013) (p. 24)

DAT204—Building Scalable Applications on AWS NoSQL Services (re:Invent 2015)

In this session, we discuss the benefits of NoSQL databases and take a tour of the main NoSQL services offered by AWS—Amazon DynamoDB and Amazon ElastiCache. Then, we hear from two leading customers, Expedia and Mapbox, about their use cases and architectural challenges, and how they addressed them using AWS NoSQL services, including design patterns and best practices. You should come out of this session having a better understanding of NoSQL and its powerful capabilities, ready to tackle your database challenges with confidence.

DAT204—Building Scalable Applications on AWS NoSQL Services (re:Invent 2015)

DAT207—Accelerating Application Performance with Amazon ElastiCache (AWS re:Invent 2013)

In this video, learn how you can use Amazon ElastiCache to easily deploy an in-memory caching system to speed up your application performance. We show you how to use Amazon ElastiCache to improve your application latency and reduce the load on your database servers. We'll also show you how to build a caching layer that is easy to manage and scale as your application grows. During this session, we go over various scenarios and use cases that can benefit by enabling caching, and discuss the features provided by Amazon ElastiCache.

DAT207 - Accelerating Application Performance with Amazon ElastiCache (re:Invent 2013)

Advanced Videos

The following videos cover more advanced Amazon ElastiCache topics.

Topics
• SDD402—Amazon ElastiCache Deep Dive (re:Invent 2014) (p. 25)
• DAT307—Deep Dive into Amazon ElastiCache Architecture and Design Patterns (re:Invent 2013) (p. 25)

DAT305—Amazon ElastiCache Deep Dive (re:Invent 2017)

Look behind the scenes to learn about Amazon ElastiCache's design and architecture. See common design patterns with our Memcached and Redis offerings and how customers have used them for in-memory operations to reduce latency and improve application throughput. During this video, we review ElastiCache best practices, design patterns, and anti-patterns.

The video introduces the following:

• ElastiCache for Redis online resharding
• ElastiCache security and encryption
• ElastiCache for Redis version 3.2.10

DAT305—Amazon ElastiCache Deep Dive (re:Invent 2017)

DAT306—Amazon ElastiCache Deep Dive (re:Invent 2016)

Look behind the scenes to learn about Amazon ElastiCache's design and architecture. See common design patterns with our Memcached and Redis offerings and how customers have used them for in-memory operations to reduce latency and improve application throughput. During this session, we review ElastiCache best practices, design patterns, and anti-patterns.

DAT306—Amazon ElastiCache Deep Dive (re:Invent 2016)

DAT407—Amazon ElastiCache Deep Dive (re:Invent 2015)

Peek behind the scenes to learn about Amazon ElastiCache's design and architecture. See common design patterns of our Memcached and Redis offerings and how customers have used them for in-memory operations and achieved improved latency and throughput for applications. During this session, we review best practices, design patterns, and anti-patterns related to Amazon ElastiCache.

DAT407—Amazon ElastiCache Deep Dive (re:Invent 2015)

SDD402—Amazon ElastiCache Deep Dive (re:Invent 2014)

In this video, we examine common caching use cases, the Memcached and Redis engines, patterns that help you determine which engine is better for your needs, consistent hashing, and more as means to building fast, scalable applications. Frank Wiebe, Principal Scientist at Adobe, details how Adobe uses Amazon ElastiCache to improve customer experience and scale their business.

SDD402—Amazon ElastiCache Deep Dive (re:Invent 2014)

DAT307—Deep Dive into Amazon ElastiCache Architecture and Design Patterns (re:Invent 2013)

In this video, we examine caching, caching strategies, scaling out, monitoring. We also compare the Memcached and Redis engines. During this session, also we review best practices and design patterns related to Amazon ElastiCache.

Caching Strategies and Best Practices

Following, you can find recommended best practices for Amazon ElastiCache. Following these improves your cluster's performance and reliability.

Topics
- Caching Strategies (p. 26)
- Configuring Your ElastiCache Client for Efficient Load Balancing (p. 31)

Caching Strategies

In the following topic, you can find strategies for populating and maintaining your cache.

What strategies to implement for populating and maintaining your cache depend upon what data you cache and the access patterns to that data. For example, you likely don't want to use the same strategy for both a top-10 leaderboard on a gaming site and trending news stories. In the rest of this section, we discuss common cache maintenance strategies and their advantages and disadvantages.

Topics
- Lazy Loading (p. 26)
- Write-Through (p. 28)
- Adding TTL (p. 29)
- Related Topics (p. 30)

Lazy Loading

As the name implies, lazy loading is a caching strategy that loads data into the cache only when necessary. It works as described following.

Amazon ElastiCache is an in-memory key-value store that sits between your application and the data store (database) that it accesses. Whenever your application requests data, it first makes the request to the ElastiCache cache. If the data exists in the cache and is current, ElastiCache returns the data to your application. If the data doesn't exist in the cache or has expired, your application requests the data from your data store. Your data store then returns the data to your application. Your application next writes the data received from the store to the cache. This way, it can be more quickly retrieved the next time it's requested.

A cache hit occurs when data is in the cache and isn't expired:

1. Your application requests data from the cache.
2. The cache returns the data to the application.

A cache miss occurs when data isn't in the cache or is expired:
1. Your application requests data from the cache.
2. The cache doesn't have the requested data, so returns a `null`.
3. Your application requests and receives the data from the database.
4. Your application updates the cache with the new data.

The following diagram illustrates both these processes.

**Advantages and Disadvantages of Lazy Loading**

The advantages of lazy loading are as follows:

- Only requested data is cached.
  
  Because most data is never requested, lazy loading avoids filling up the cache with data that isn't requested.

- Node failures aren't fatal for your application.
  
  When a node fails and is replaced by a new, empty node, your application continues to function, though with increased latency. As requests are made to the new node, each cache miss results in a query of the database. At the same time, the data copy is added to the cache so that subsequent requests are retrieved from the cache.

The disadvantages of lazy loading are as follows:

- There is a cache miss penalty. Each cache miss results in three trips:
  
  1. Initial request for data from the cache
  2. Query of the database for the data
  3. Writing the data to the cache
  
  These misses can cause a noticeable delay in data getting to the application.

- Stale data.
  
  If data is written to the cache only when there is a cache miss, data in the cache can become stale. This result occurs because there are no updates to the cache when data is changed in the database. To address this issue, you can use the Write-Through (p. 28) and Adding TTL (p. 29) strategies.

**Lazy Loading Pseudocode Example**

The following is a pseudocode example of lazy loading logic.
For this example, the application code that gets the data is the following.

```java
customer_record = get_customer(12345)
```

**Write-Through**

The write-through strategy adds data or updates data in the cache whenever data is written to the database.

**Advantages and Disadvantages of Write-Through**

The advantages of write-through are as follows:

- Data in the cache is never stale.
  
  Because the data in the cache is updated every time it's written to the database, the data in the cache is always current.

- Write penalty vs. read penalty.
  
  Every write involves two trips:
  1. A write to the cache
  2. A write to the database

  Which adds latency to the process. That said, end users are generally more tolerant of latency when updating data than when retrieving data. There is an inherent sense that updates are more work and thus take longer.

The disadvantages of write-through are as follows:

- Missing data.
  
  If you spin up a new node, whether due to a node failure or scaling out, there is missing data. This data continues to be missing until it's added or updated on the database. You can minimize this by implementing lazy loading (p. 26) with write-through.

- Cache churn.
  
  Most data is never read, which is a waste of resources. By adding a time to live (TTL) value (p. 29), you can minimize wasted space.
Write-Through Pseudocode Example

The following is a pseudocode example of write-through logic.

```java
// *****************************************
// function that saves a customer's record.
// *****************************************
save_customer(customer_id, values)
    customer_record = db.query("UPDATE Customers WHERE id = {0}\n" "value", customer_id, values)
    cache.set(customer_id, customer_record)
    return success
```

For this example, the application code that gets the data is the following.

```
save_customer(12345,{"address":"123 Main"})
```

Adding TTL

Lazy loading allows for stale data but doesn’t fail with empty nodes. Write-through ensures that data is always fresh, but can fail with empty nodes and can populate the cache with superfluous data. By adding a time to live (TTL) value to each write, you can have the advantages of each strategy. At the same time, you can largely avoid cluttering up the cache with extra data.

Time to live (TTL) is an integer value that specifies the number of seconds until the key expires. Redis can specify seconds or milliseconds for this value. When an application attempts to read an expired key, it is treated as though the key is not found. The database is queried for the key and the cache is updated. This approach doesn’t guarantee that a value isn’t stale. However, it keeps data from getting too stale and requires that values in the cache are occasionally refreshed from the database.

For more information, see the Memcached `set` command.

TTL Pseudocode Examples

The following is a pseudocode example of write-through logic with TTL.

```java
// *****************************************
// function that saves a customer's record.
// The TTL value of 300 means that the record expires
// 300 seconds (5 minutes) after the set command
// and future reads will have to query the database.
// *****************************************
save_customer(customer_id, values)
    customer_record = db.query("UPDATE Customers WHERE id = {0}\n" "value", customer_id, values)
    cache.set(customer_id, customer_record, 300)
    return success
```

The following is a pseudocode example of lazy loading logic with TTL.

```java
```

// *****************************************
// function that returns a customer's record.
// Attempts to retrieve the record from the cache.
// If it is retrieved, the record is returned to the application.
// If the record is not retrieved from the cache, it is
// retrieved from the database,
// added to the cache, and
// returned to the application.
// The TTL value of 300 means that the record expires
// 300 seconds (5 minutes) after the set command
// and subsequent reads will have to query the database.
// *****************************************

get_customer(customer_id)

customer_record = cache.get(customer_id)

if (customer_record != null)
    if (customer_record.TTL < 300)
        return customer_record        // return the record and exit function

    // do this only if the record did not exist in the cache OR
    // the TTL was >= 300, i.e., the record in the cache had expired.
    customer_record = db.query("SELECT * FROM Customers WHERE id = {0} ", customer_id)
    cache.set(customer_id, customer_record, 300)  // update the cache
    return customer_record        // return the newly retrieved record and exit
function

For this example, the application code that gets the data is the following.

save_customer(12345, {"address": "123 Main"})

customer_record = get_customer(12345)

Related Topics

- In-Memory Data Store (p. 2)
- Choosing an Engine and Version
- Scaling ElastiCache for Memcached Clusters (p. 132)
Configuring Your ElastiCache Client for Efficient Load Balancing

**Note**  
This section applies to multi-node Memcached clusters.

To effectively use multiple ElastiCache Memcached nodes, you need to be able to spread your cache keys across the nodes. A simple way to load balance a cluster with $n$ nodes is to calculate the hash of the object's key and mod the result by $n$ - hash(key) mod $n$. The resulting value (0 through $n-1$) is the number of the node where you place the object.

This approach is simple and works well as long as the number of nodes ($n$) is constant. However, whenever you add or remove a node from the cluster, the number of keys that need to be moved is $(n - 1) / n$ (where $n$ is the new number of nodes). Thus, this approach results in a large number of keys being moved, which translates to a large number of initial cache misses, especially as the number of nodes gets large. Scaling from 1 to 2 nodes results in $(2-1) / 2$ (50 percent) of the keys being moved, the best case. Scaling from 9 to 10 nodes results in $(10–1)/10$ (90 percent) of the keys being moved. If you're scaling up due to a spike in traffic, you don't want to have a large number of cache misses. A large number of cache misses results in hits to the database, which is already overloaded due to the spike in traffic.

The solution to this dilemma is consistent hashing. Consistent hashing uses an algorithm such that whenever a node is added or removed from a cluster, the number of keys that must be moved is roughly $1 / n$ (where $n$ is the new number of nodes). Scaling from 1 to 2 nodes results in $1/2$ (50 percent) of the keys being moved, the worst case. Scaling from 9 to 10 nodes results in $1/10$ (10 percent) of the keys being moved.

As the user, you control which hashing algorithm is used for multi-node clusters. We recommend that you configure your clients to use consistent hashing. Fortunately, there are many Memcached client libraries in most popular languages that implement consistent hashing. Check the documentation for the library you are using to see if it supports consistent hashing and how to implement it.

If you are working in Java, PHP, or .NET, we recommend you use one of the Amazon ElastiCache client libraries.

### Consistent Hashing Using Java

The ElastiCache Memcached Java client is based on the open-source spymemcached Java client, which has consistent hashing capabilities built in. The library includes a KetamaConnectionFactory class that implements consistent hashing. By default, consistent hashing is turned off in spymemcached.

For more information, see the KetamaConnectionFactory documentation at http://dustin.sallings.org/java-memcached-client/apidocs/net/spy/memcached/KetamaConnectionFactory.html.

### Consistent Hashing Using PHP

The ElastiCache Memcached PHP client is a wrapper around the built-in Memcached PHP library. By default, consistent hashing is turned off by the Memcached PHP library.

Use the following code to turn on consistent hashing.

```php
$m = new Memcached();
$m->setOption(Memcached::OPT_DISTRIBUTION, Memcached::DISTRIBUTION_CONSISTENT);
```

In addition to the preceding code, we recommend that you also turn `memcached.sessconsistent_hash` on in your `php.ini` file.
For more information, see the run-time configuration documentation for Memcached PHP at http://php.net/manual/en/memcached.configuration.php. Note specifically the memcached.sess_consistent_hash parameter.

**Consistent Hashing Using .NET**

The ElastiCache Memcached .NET client is a wrapper around Enyim Memcached. By default, consistent hashing is turned on by the Enyim Memcached client.

For more information, see the memcached/locator documentation at https://github.com/enyim/EnyimMemcached/wiki/MemcachedClient-Configuration#user-content-memcachedlocator.
Managing Your ElastiCache for Memcached Implementation

In this section, you can find details about how to manage the various components of your ElastiCache implementation. These include tasks such as creating, updating, and deleting nodes or clusters, and many more.

Topics
- Engine Versions and Upgrading (p. 33)
- Choosing Regions and Availability Zones (p. 37)
- Managing Maintenance (p. 41)
- Managing Nodes (p. 42)
- Managing Your ElastiCache Clusters (p. 51)
- Accessing Your Cluster (p. 107)
- Replication Across AWS Regions Using Global Datastore (p. 115)
- Finding Connection Endpoints (p. 126)
- Scaling ElastiCache for Memcached Clusters (p. 132)
- Configuring Engine Parameters Using Parameter Groups (p. 135)

Engine Versions and Upgrading

This section covers the supported Memcached engine versions and how to upgrade.

Topics
- Supported ElastiCache for Memcached Versions (p. 34)
- Upgrading Engine Versions (p. 36)
Supported ElastiCache for Memcached Versions

ElastiCache supports the following Memcached versions and upgrading to newer versions. When upgrading to a newer version, pay careful attention to the conditions that if not met cause your upgrade to fail.

ElastiCache for Memcached Versions

- Memcached Version 1.5.16 (p. 34)
- Memcached Version 1.5.10 (p. 34)
- Memcached Version 1.4.34 (p. 34)
- Memcached Version 1.4.33 (p. 35)
- Memcached Version 1.4.24 (p. 35)
- Memcached Version 1.4.14 (p. 35)
- Memcached Version 1.4.5 (p. 35)

Memcached Version 1.5.16

ElastiCache for Memcached adds support for Memcached version 1.5.16. It includes no new features, but does include bug fixes and cumulative updates from Memcached 1.5.14 and Memcached 1.5.15.

For more information, see Memcached 1.5.16 Release Notes at Memcached on GitHub.

Memcached Version 1.5.10

ElastiCache for Memcached version 1.5.10 supports the following Memcached features:

- Automated slab rebalancing.
- Faster hash table lookups with murmur3 algorithm.
- Segmented LRU algorithm.
- LRU crawler to background-reclaim memory.
- --enable-seccomp: A compile-time option.

It also introduces the no_modern and inline_ascii_resp parameters. For more information, see Memcached 1.5.10 Parameter Changes (p. 151).

Memcached improvements added since ElastiCache for Memcached version 1.4.34 include the following:

- Cumulative fixes, such as ASCII multigets, CVE-2017-9951 and limit crawls for metadumper.
- Better connection management by closing connections at the connection limit.
- Improved item-size management for item size above 1MB.
- Better performance and memory-overhead improvements by reducing memory requirements per-item by a few bytes.

For more information, see Memcached 1.5.10 Release Notes at Memcached on GitHub.

Memcached Version 1.4.34

ElastiCache for Memcached version 1.4.34 adds no new features to version 1.4.33. Version 1.4.34 is a bug fix release that is larger than the usual such release.

For more information, see Memcached 1.4.34 Release Notes at Memcached on GitHub.
Memcached Version 1.4.33

Memcached improvements added since version 1.4.24 include the following:

- Ability to dump all of the metadata for a particular slab class, a list of slab classes, or all slab classes. For more information, see Memcached 1.4.31 Release Notes.
- Improved support for large items over the 1 megabyte default. For more information, see Memcached 1.4.29 Release Notes.
- Ability to specify how long a client can be idle before being asked to close.

Ability to dynamically increase the amount of memory available to Memcached without having to restart the cluster. For more information, see Memcached 1.4.27 Release Notes.

- Logging of fetchers, mutations, and evictions are now supported. For more information, see Memcached 1.4.26 Release Notes.
- Freed memory can be reclaimed back into a global pool and reassigned to new slab classes. For more information, see Memcached 1.4.25 Release Notes.
- Several bug fixes.
- Some new commands and parameters. For a list, see Memcached 1.4.33 Added Parameters (p. 152).

Memcached Version 1.4.24

Memcached improvements added since version 1.4.14 include the following:

- Least recently used (LRU) management using a background process.
- Added the option of using jenkins or murmur3 as your hash algorithm.
- Some new commands and parameters. For a list, see Memcached 1.4.24 Added Parameters (p. 154).
- Several bug fixes.

Memcached Version 1.4.14

Memcached improvements added since version 1.4.5 include the following:

- Enhanced slab rebalancing capability.
- Performance and scalability improvement.
- Introduced the touch command to update the expiration time of an existing item without fetching it.
- Auto discovery—the ability for client programs to automatically determine all of the cache nodes in a cluster, and to initiate and maintain connections to all of these nodes.

Memcached Version 1.4.5

Memcached version 1.4.5 was the initial engine and version supported by Amazon ElastiCache for Memcached.
Upgrading Engine Versions

You can control if and when the protocol-compliant software powering your cache cluster is upgraded to new versions that are supported by ElastiCache. This level of control enables you to maintain compatibility with specific versions, test new versions with your application before deploying in production, and perform version upgrades on your own terms and timelines.

Because version upgrades might involve some compatibility risk, they don't occur automatically. You must initiate them.

To upgrade to a newer Memcached version, modify your cache cluster specifying the new engine version you want to use. Upgrading to a newer Memcached version is a destructive process – you lose your data and start with a cold cache. For more information, see Modifying an ElastiCache Cluster (p. 91).

You should be aware of the following requirements when upgrading from an older version of Memcached to Memcached version 1.4.33 or newer. CreateCacheCluster and ModifyCacheCluster fails under the following conditions:

- If $\text{slab\_chunk\_max} > \text{max\_item\_size}$.
- If $\text{max\_item\_size} \mod \text{slab\_chunk\_max} \neq 0$.
- If $\text{max\_item\_size} > (\text{max\_cache\_memory} - \text{memcached\_connections\_overhead}) / 4$.

The value $(\text{max\_cache\_memory} - \text{memcached\_connections\_overhead})$ is the node's memory useable for data. For more information, see Memcached Connection Overhead (p. 158).

**Important**

- You can upgrade to a newer engine version, but you can't downgrade to an older engine version. If you want to use an older engine version, you must delete the existing cluster and create it anew with the older engine version.
- Engine version management is designed so that you can have as much control as possible over how patching occurs. However, ElastiCache reserves the right to patch your cluster on your behalf in the unlikely event of a critical security vulnerability in the system or cache software.
- Starting with Redis engine version 5.0.5, you can upgrade your cluster version with minimal downtime. The cluster is available for reads during the entire upgrade and is available for writes for most of the upgrade duration, except during the failover operation which lasts a few seconds.
- You can also upgrade your ElastiCache clusters with versions earlier than 5.0.5. The process involved is the same but may incur longer failover time during DNS propagation (30s-1m).
- Because the Memcached engine does not support persistence, Memcached engine version upgrades are always a disruptive process that clears all cache data in the cluster.

How to Upgrade Engine Versions

To start version upgrades to your cluster, you modify it and specify a newer engine version. You can do this by using the ElastiCache console, the AWS CLI, or the ElastiCache API:

- To use the AWS Management Console, see – Using the AWS Management Console (p. 91).
- To use the AWS CLI, see Using the AWS CLI (p. 91).
- To use the ElastiCache API, see Using the ElastiCache API (p. 92).
Choosing Regions and Availability Zones

AWS Cloud computing resources are housed in highly available data center facilities. To provide additional scalability and reliability, these data center facilities are located in different physical locations. These locations are categorized by regions and Availability Zones.

AWS Regions are large and widely dispersed into separate geographic locations. Availability Zones are distinct locations within an AWS Region that are engineered to be isolated from failures in other Availability Zones. They provide inexpensive, low-latency network connectivity to other Availability Zones in the same AWS Region.

**Important**
Each region is completely independent. Any ElastiCache activity you initiate (for example, creating clusters) runs only in your current default region.

To create or work with a cluster in a specific region, use the corresponding regional service endpoint. For service endpoints, see Supported Regions & Endpoints (p. 38).

---

**Regions and Availability Zones**

**Topics**
- Supported Regions & Endpoints (p. 38)
- Locating Your Nodes (p. 41)
Supported Regions & Endpoints

Amazon ElastiCache is available in multiple AWS Regions. This means that you can launch ElastiCache clusters in locations that meet your requirements. For example, you can launch in the AWS Region closest to your customers, or launch in a particular AWS Region to meet certain legal requirements.

By default, the AWS SDKs, AWS CLI, ElastiCache API, and ElastiCache console reference the US-West (Oregon) region. As ElastiCache expands availability to new regions, new endpoints for these regions are also available to use in your HTTP requests, the AWS SDKs, AWS CLI, and the console.

Each region is designed to be completely isolated from the other regions. Within each region are multiple Availability Zones (AZ). By launching your nodes in different AZs you are able to achieve the greatest possible fault tolerance. For more information on regions and Availability Zones, see Choosing Regions and Availability Zones (p. 37) at the top of this topic.

Regions where ElastiCache is supported

<table>
<thead>
<tr>
<th>Region Name/Region</th>
<th>Endpoint</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>US East (Ohio) Region</td>
<td>elasticache.us-east-2.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>us-east-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US East (N. Virginia) Region</td>
<td>elasticache.us-east-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>us-east-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US West (N. California) Region</td>
<td>elasticache.us-west-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>us-west-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US West (Oregon) Region</td>
<td>elasticache.us-west-2.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>us-west-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada (Central) Region</td>
<td>elasticache.ca-central-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ca-central-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific (Mumbai) Region</td>
<td>elasticache.ap-south-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ap-south-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific (Tokyo) Region</td>
<td>elasticache.ap-northeast-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ap-northeast-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific (Seoul) Region</td>
<td>elasticache.ap-northeast-2.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ap-northeast-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific (Osaka-Local) Region *</td>
<td>elasticache.ap-northeast-3.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ap-northeast-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region Name/Region</td>
<td>Endpoint</td>
<td>Protocol</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Asia Pacific (Singapore) Region</td>
<td>elasticache.ap-southeast-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ap-southeast-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific (Sydney) Region</td>
<td>elasticache.ap-southeast-2.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ap-southeast-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe (Frankfurt) Region</td>
<td>elasticache.eu-central-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>eu-central-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe (Ireland) Region</td>
<td>elasticache.eu-west-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>eu-west-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe (London) Region</td>
<td>elasticache.eu-west-2.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>eu-west-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU (Paris) Region</td>
<td>elasticache.eu-west-3.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>eu-west-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America (São Paulo) Region</td>
<td>elasticache.sa-east-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>sa-east-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China (Beijing) Region</td>
<td>elasticache.cn-north-1.amazonaws.com.cn</td>
<td>HTTPS</td>
</tr>
<tr>
<td>cn-north-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>China (Ningxia) Region</td>
<td>elasticache.cn-northwest-1.amazonaws.com.cn</td>
<td>HTTPS</td>
</tr>
<tr>
<td>cn-northwest-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific (Hong Kong) Region</td>
<td>elasticache.ap-east-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>ap-east-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AWS GovCloud (US-West)</td>
<td>elasticache.us-gov-west-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>us-gov-west-1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For information on using the AWS GovCloud (US) with ElastiCache, see Services in the AWS GovCloud (US) region: ElastiCache.

Notes:
The Asia Pacific (Osaka-Local) Region is a local region that is available to select AWS customers who request access. If you want to use the Asia Pacific (Osaka-Local) Region, speak with your sales representative. The Asia Pacific (Osaka-Local) Region supports a single Availability Zone.
Some regions support a subset of node types. For a table of supported node types by AWS Region, see Supported Node Types by AWS Region (p. 47).

For a table of AWS products and services by region, see Products and Services by Region.
Locating Your Nodes

Amazon ElastiCache supports locating all of a cluster's nodes in a single or multiple Availability Zones (AZs). Further, if you elect to locate your nodes in multiple AZs (recommended), ElastiCache enables you to either choose the AZ for each node, or allow ElastiCache to choose them for you.

By locating the nodes in different AZs, you eliminate the chance that a failure, such as a power outage, in one AZ will cause your entire system to fail. Testing has demonstrated that there is no significant latency difference between locating all nodes in one AZ or spreading them across multiple AZs.

You can specify an AZ for each node when you create a cluster or by adding nodes when you modify an existing cluster. For more information, see the following:

- Creating a Cluster (p. 53)
- Modifying an ElastiCache Cluster (p. 91)
- Adding Nodes to a Cluster (p. 94)

Managing Maintenance

Every cluster has a weekly maintenance window during which any system changes are applied. If you don't specify a preferred maintenance window when you create or modify a cluster, ElastiCache assigns a 60-minute maintenance window within your region's maintenance window on a randomly chosen day of the week.

The 60-minute maintenance window is chosen at random from an 8-hour block of time per region. The following table lists the time blocks for each region from which the default maintenance windows are assigned. You may choose a preferred maintenance window outside the region's maintenance window block.

<table>
<thead>
<tr>
<th>Region Code</th>
<th>Region Name</th>
<th>Region Maintenance Window</th>
</tr>
</thead>
<tbody>
<tr>
<td>ap-northeast-1</td>
<td>Asia Pacific (Tokyo) Region</td>
<td>13:00–21:00 UTC</td>
</tr>
<tr>
<td>ap-northeast-2</td>
<td>Asia Pacific (Seoul) Region</td>
<td>12:00–20:00 UTC</td>
</tr>
<tr>
<td>ap-northeast-3</td>
<td>Asia Pacific (Osaka-Local) Region</td>
<td>12:00–20:00 UTC</td>
</tr>
<tr>
<td>ap-south-1</td>
<td>Asia Pacific (Mumbai) Region</td>
<td>17:30–1:30 UTC</td>
</tr>
<tr>
<td>ap-southeast-1</td>
<td>Asia Pacific (Singapore) Region</td>
<td>14:00–22:00 UTC</td>
</tr>
<tr>
<td>ap-southeast-2</td>
<td>Asia Pacific (Sydney) Region</td>
<td>12:00–20:00 UTC</td>
</tr>
<tr>
<td>cn-north-1</td>
<td>China (Beijing) region</td>
<td>14:00–22:00 UTC</td>
</tr>
<tr>
<td>eu-central-1</td>
<td>Europe (Frankfurt) Region</td>
<td>23:00–07:00 UTC</td>
</tr>
<tr>
<td>eu-west-1</td>
<td>Europe (Ireland) Region</td>
<td>22:00–06:00 UTC</td>
</tr>
<tr>
<td>eu-west-2</td>
<td>Europe (London) Region</td>
<td>23:00–07:00 UTC</td>
</tr>
<tr>
<td>sa-east-1</td>
<td>South America (São Paulo) Region</td>
<td>01:00–09:00 UTC</td>
</tr>
<tr>
<td>us-east-1</td>
<td>US East (N. Virginia) Region</td>
<td>03:00–11:00 UTC</td>
</tr>
<tr>
<td>us-east-2</td>
<td>US East (Ohio) Region</td>
<td>04:00–12:00 UTC</td>
</tr>
</tbody>
</table>

API Version 2015-02-02

41
### Managing Nodes

A node is the smallest building block of an Amazon ElastiCache deployment. It is a fixed-size chunk of secure, network-attached RAM. Each node runs the engine that was chosen when the cluster was created or last modified. Each node has its own Domain Name Service (DNS) name and port. Multiple types of ElastiCache nodes are supported, each with varying amounts of associated memory and computational power.

Generally speaking, due to its support for sharding, Memcached deployments have a number of smaller nodes. For a more detailed discussion of which node size to use, see Choosing Your Memcached Node Size (p. 80).

**Topcs**
- [Connecting to Nodes](p. 43)
- [ElastiCache Reserved Nodes](p. 45)
- [Supported Node Types](p. 46)
- [Replacing Nodes](p. 50)

Some important operations involving nodes are the following:
- [Adding Nodes to a Cluster](p. 94)
- [Removing Nodes from a Cluster](p. 99)
- [Scaling ElastiCache for Memcached Clusters](p. 132)
- [Finding Connection Endpoints](p. 126)
- [Automatically Identify Nodes in your Memcached Cluster](p. 55)
Connecting to Nodes

Before attempting to connect to your Memcached cluster, you must have the endpoints for the nodes. To find the endpoints, see the following:

- Finding a Cluster’s Endpoints (Console) (p. 128)
- Finding Endpoints (AWS CLI) (p. 130)
- Finding Endpoints (ElastiCache API) (p. 132)

In the following example, you use the `telnet` utility to connect to a node that is running Memcached.

**Note**
For more information about Memcached and available Memcached commands, see the Memcached website.

**To connect to a node using `telnet`**

1. Connect to your Amazon EC2 instance by using the connection utility of your choice.
   
   **Note**
   For instructions on how to connect to an Amazon EC2 instance, see the Amazon EC2 Getting Started Guide.

2. Download and install the `telnet` utility on your Amazon EC2 instance. At the command prompt of your Amazon EC2 instance, type the following command and type `y` at the command prompt.

   ```
   sudo yum install telnet
   ```

   Output similar to the following appears.

   ```
   Loaded plugins: priorities, security, update-motd, upgrade-helper
   Setting up Install Process
   Resolving Dependencies
   --> Running transaction check
   ...(output omitted)...
   Total download size: 63 k
   Installed size: 109 k
   Is this ok [y/N]: y
   Downloading Packages:
   telnet-0.17-47.7.amzn1.x86_64.rpm | 63 kB 00:00
   ...(output omitted)...
   Complete!
   ```

3. At the command prompt of your Amazon EC2 instance, type the following command, substituting the endpoint of your node for the one shown in this example.

   ```
   telnet mycachecluster.eaogs8.0001.usw2.cache.amazonaws.com 11211
   ```

   Output similar to the following appears.

   ```
   Trying 128.0.0.1...
   Connected to mycachecluster.eaogs8.0001.usw2.cache.amazonaws.com.
   ```
4. Test the connection by running Memcached commands.

You are now connected to a node, and you can run Memcached commands. The following is an example.

```
set a 0 0 5     // Set key "a" with no expiration and 5 byte value
hello          // Set value as "hello"
STORED
get a           // Get value for key "a"
VALUE a 0 5   hello
END
get b           // Get value for key "b" results in miss
END
```

Escape character is '^]'.
>
ElastiCache Reserved Nodes

Reserving one or more nodes might be a way for you to reduce costs. Reserved nodes are charged an up front fee that depends upon the node type and the length of reservation— one or three years. In addition to the upfront charge, there is an hourly usage charge. This charge is much less than the hourly usage charge you incur with On-Demand nodes.

To see if reserved nodes are a cost savings for your use cases, first determine the node size and number of nodes you need. Then estimate the usage of the node, and compare the total cost to you of using On-Demand nodes versus reserved nodes. You can mix and match reserved and On-Demand node usage in your clusters. For pricing information, see Amazon ElastiCache Pricing.

For more information, see Managing Costs with Reserved Nodes (p. 186).
Supported Node Types

ElastiCache supports the following node types. Generally speaking, the current generation types provide more memory and computational power at lower cost when compared to their equivalent previous generation counterparts.

- **General purpose:**
  - Current generation:

    - **M5 node types:** cache.m5.large, cache.m5.xlarge, cache.m5.2xlarge, cache.m5.4xlarge, cache.m5.12xlarge, cache.m5.24xlarge

    - **M4 node types:** cache.m4.large, cache.m4.xlarge, cache.m4.2xlarge, cache.m4.4xlarge, cache.m4.10xlarge

    - **T3 node types:** cache.t3.micro, cache.t3.small, cache.t3.medium

    - **T2 node types:** cache.t2.micro, cache.t2.small, cache.t2.medium

  - Previous generation: (not recommended)

    - **T1 node types:** cache.t1.micro

    - **M1 node types:** cache.m1.small, cache.m1.medium, cache.m1.large, cache.m1.xlarge

    - **M3 node types:** cache.m3.medium, cache.m3.large, cache.m3.xlarge, cache.m3.2xlarge

  - **Compute optimized:**

    - Previous generation: (not recommended)

    - **C1 node types:** cache.c1.xlarge

  - **Memory optimized:**

    - Current generation:

    - **R5 node types:** cache.r5.large, cache.r5.xlarge, cache.r5.2xlarge, cache.r5.4xlarge, cache.r5.12xlarge, cache.r5.24xlarge

    - **R4 node types:** cache.r4.large, cache.r4.xlarge, cache.r4.2xlarge, cache.r4.4xlarge, cache.r4.8xlarge, cache.r4.16xlarge

    - Previous generation: (not recommended)

    - **M2 node types:** cache.m2.xlarge, cache.m2.2xlarge, cache.m2.4xlarge

    - **R3 node types:** cache.r3.large, cache.r3.xlarge, cache.r3.2xlarge, cache.r3.4xlarge, cache.r3.8xlarge

You can launch general-purpose burstable T3-Standard cache nodes in Amazon ElastiCache. These nodes provide a baseline level of CPU performance with the ability to burst CPU usage at any time until the accrued credits are exhausted. A CPU credit provides the performance of a full CPU core for one minute.

Amazon ElastiCache's T3 nodes are configured as standard and suited for workloads with an average CPU utilization that is consistently below the baseline performance of the instance. To burst above the baseline, the node spends credits that it has accrued in its CPU credit balance. If the node is running low on accrued credits, performance is gradually lowered to the baseline performance level. This gradual lowering ensures the node doesn't 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 in the Amazon EC2 User Guide.
The following table lists the burstable performance node types, the rate at which CPU credits are earned per hour. It also shows the maximum number of earned CPU credits that a node can accrue and the number of vCPUs per node. In addition, it gives the baseline performance level as a percentage of a full core performance (using a single vCPU).

<table>
<thead>
<tr>
<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>
<th>Memory (GiB)</th>
<th>Network Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3.micro</td>
<td>288</td>
<td>2</td>
<td>10%</td>
<td>0.5</td>
<td>Up to 5 Gigabit</td>
</tr>
<tr>
<td>t3.small</td>
<td>576</td>
<td>2</td>
<td>20%</td>
<td>1.37</td>
<td>Up to 5 Gigabit</td>
</tr>
<tr>
<td>t3.medium</td>
<td>576</td>
<td>2</td>
<td>20%</td>
<td>3.09</td>
<td>Up to 5 Gigabit</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. Some node sizes that have more than one vCPU. For these, calculate the baseline CPU utilization for the node by multiplying the vCPU percentage by the number of vCPUs.

The following CPU credit metrics are available for burstable performance instances:

- CPUCreditUsage
- CPUCreditBalance

For more information on these metrics, see CPU Credit Metrics.

In addition, be aware of these details:

- All current generation node types are created in a virtual private cloud (VPC) based on Amazon VPC by default.

** Note **

Supported engine versions vary by AWS Region. The latest engine versions are supported in all AWS Regions. To find the available engine versions in your AWS Region, see Supported ElastiCache for Memcached Versions (p. 34).

Supported Node Types by AWS Region

The following table lists supported node types for each AWS Region.

<table>
<thead>
<tr>
<th>AWS Region</th>
<th>T3</th>
<th>T2</th>
<th>M4</th>
<th>M5</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-east-2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>East (Ohio)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

API Version 2015-02-02

47
<table>
<thead>
<tr>
<th>AWS Region</th>
<th>T3</th>
<th>T2</th>
<th>M4</th>
<th>M5</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-east-1 East (N. Virginia)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>US-west-1 West (N. California)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>US-west-2 West (Oregon)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Canada-central-1 (Central)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Asia-south-1 Pacific (Mumbai)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Asia-northeast-1 Pacific (Tokyo)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Asia-northeast-2 Pacific (Seoul)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Asia-northeast-3 Pacific (Osaka-Local) *</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Asia-southeast-1 Pacific (Singapore)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Asia-southeast-2 Pacific (Sydney)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Asia-east-1 Pacific (Hong Kong)</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Europe-1 (Stockholm)</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Europe-central-1 (Frankfurt)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Europe-1 (Ireland)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Europe-2 (London)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
### Supported Node Types

<table>
<thead>
<tr>
<th>AWS Region</th>
<th>T3</th>
<th>T2</th>
<th>M4</th>
<th>M5</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU-west-3 (Paris)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>South America (São Paulo)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>China north-1 (Beijing)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>China northwest-1 (Ningxia)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Middle East (Bahrain)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AWS gov-west-1 (US-West)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* The Asia Pacific (Osaka-Local) Region is a local region that is available to select AWS customers who request access. If you want to use the Asia Pacific (Osaka-Local) Region, speak with your sales representative. The Asia Pacific (Osaka-Local) Region supports a single Availability Zone.

For a complete list of node types and specifications, see the following:

- Amazon ElastiCache Product Features and Details
- Memcached Node-Type Specific Parameters
Replacing Nodes

Amazon ElastiCache for Memcached frequently upgrades its fleet with patches and upgrades being applied to instances seamlessly. However, from time to time we need to relaunch your ElastiCache for Memcached nodes to apply mandatory OS updates to the underlying host. These replacements are required to apply upgrades that strengthen security, reliability, and operational performance.

You have the option to manage these replacements yourself at any time before the scheduled node replacement window. When you manage a replacement yourself, your instance receives the OS update when you relaunch the node and your scheduled node replacement is canceled. You might continue to receive alerts indicating that the node replacement takes place. If you already manually mitigated the need for the maintenance, you can ignore these alerts.

**Note**
Replacement cache nodes automatically generated by Amazon ElastiCache may have different IP addresses. You are responsible for reviewing your application configuration to ensure that your cache nodes are associated with the appropriate IP addresses.

The following list identifies actions you can take when ElastiCache schedules one of your Memcached nodes for replacement.

- **Do nothing** – If you do nothing, ElastiCache replaces the node as scheduled. When ElastiCache automatically replaces the node with a new node, the new node is initially empty.

- **Change your maintenance window** – For scheduled maintenance events, you receive an email or a notification event from ElastiCache. In this case, if you change your maintenance window before the scheduled replacement time, your node now is replaced at the new time. For more information, see [Modifying an ElastiCache Cluster](p. 91).

  **Note**
  The ability to change your replacement window by moving your maintenance window is only available when the ElastiCache notification includes a maintenance window. If the notification does not include a maintenance window, you cannot change your replacement window.

For example, let’s say it’s Thursday, November 9, at 15:00 and the next maintenance window is Friday, November 10, at 17:00. Following are three scenarios with their outcomes:

- You change your maintenance window to Fridays at 16:00, after the current date and time and before the next scheduled maintenance window. The node is replaced on Friday, November 10, at 16:00.

- You change your maintenance window to Saturday at 16:00, after the current date and time and after the next scheduled maintenance window. The node is replaced on Saturday, November 11, at 16:00.

- You change your maintenance window to Wednesday at 16:00, earlier in the week than the current date and time. The node is replaced next Wednesday, November 15, at 16:00.

For instructions, see [Managing Maintenance](p. 41).

- **Manually replace the node** – If you need to replace the node before the next maintenance window, manually replace the node.

If you manually replace the node, keys are redistributed. This redistribution causes cache misses.

**To manually replace a Memcached node**

1. Delete the node scheduled for replacement. For instructions, see [Removing Nodes from a Cluster](p. 99).

2. Add a new node to the cluster. For instructions, see [Adding Nodes to a Cluster](p. 94).
Managing Your ElastiCache Clusters

A *cluster* is a collection of one or more cache nodes, all of which run an instance of the Memcached cache engine software. When you create a cluster, you specify the engine and version that all of the nodes will use.

The following diagram illustrates a typical Memcached cluster. Memcached clusters contain from 1 to 20 nodes across which you horizontally partition your data.

![Typical Memcached Cluster](image)

**Typical Memcached Cluster**

Most ElastiCache operations are performed at the cluster level. You can set up a cluster with a specific number of nodes and a parameter group that controls the properties for each node. All nodes within a cluster are designed to be of the same node type and have the same parameter and security group settings.

Every cluster must have a cluster identifier. The cluster identifier is a customer-supplied name for the cluster. This identifier specifies a particular cluster when interacting with the ElastiCache API and AWS CLI commands. The cluster identifier must be unique for that customer in an AWS Region.

ElastiCache supports multiple engine versions. Unless you have specific reasons, we recommend always using the your engine's latest version.

ElastiCache clusters are designed to be accessed by using an Amazon EC2 instance. If you launch your cluster in a VPC based on the Amazon VPC service, you can access it from outside AWS. For more information, see the following:

- Step 2: Authorize Access (p. 19)
- Accessing ElastiCache Resources from Outside AWS (p. 111)

**Supported Memcached Versions**

- Memcached Version 1.4.34 (p. 34)
- Memcached Version 1.4.33 (p. 35)
- Memcached Version 1.4.24 (p. 35)
- Memcached Version 1.4.14 (p. 35)
- Memcached Version 1.4.5 (p. 35)

**Other ElastiCache Cluster Operations**

Additional operations involving clusters:
Other ElastiCache Cluster Operations

• Finding Connection Endpoints (p. 126)
• Accessing ElastiCache Resources from Outside AWS (p. 111)
Creating a Cluster

In this section you will find instructions on creating a cluster using the ElastiCache console, the AWS CLI, or the ElastiCache API.

You can also create an ElastiCache cluster using AWS CloudFormation. For more information, see AWS::ElastiCache::CacheCluster in the AWS Cloud Formation User Guide, which includes guidance on how to implement that approach.

Whenever you create a cluster, it is a good idea to do some preparatory work so you won’t need to upgrade or make changes right away.

Topics
- Determine Your Requirements (p. 53)
- Automatically Identify Nodes in your Memcached Cluster (p. 55)
- Choosing Your Node Size (p. 80)
- Creating a Memcached Cluster (Console) (p. 82)
- Creating a Cluster (AWS CLI) (p. 84)
- Creating a Cluster (ElastiCache API) (p. 85)

Determine Your Requirements

Topics
- Memory and Processor Requirements (p. 54)
- Memcached Cluster Configuration (p. 54)
- Scaling Requirements (p. 54)
- Access Requirements (p. 54)
- Region and Availability Zone Requirements (p. 55)

Preparation

Knowing the answers to these questions before you begin will expedite creating your cluster.

- Which node instance type do you need?

  For guidance on choosing an instance node type, see Choosing Your Memcached Node Size (p. 80).

- Will you launch your cluster in a VPC or an Amazon VPC?
  
    Important
    
    If you're going to launch your cluster in an Amazon VPC, you need to create a subnet group in the same VPC before you start creating a cluster. For more information, see Subnets and Subnet Groups (p. 214).
    
    ElastiCache is designed to be accessed from within AWS using Amazon EC2. However, if you launch in a VPC based on Amazon VPC and your cluster is in an VPC, you can provide access from outside AWS. For more information, see Accessing ElastiCache Resources from Outside AWS (p. 111).

- Do you need to customize any parameter values?

  If you do, you need to create a custom Parameter Group. For more information, see Creating a Parameter Group (p. 137).

- Do you need to create your own Security Group or VPC Security Group?

  For more information, see Security Groups: EC2-Classic (p. 222) and Security in Your VPC.
• How do you intend to implement fault tolerance?

For more information, see Mitigating Failures (p. 251).

Topics
• Memory and Processor Requirements (p. 54)
• Memcached Cluster Configuration (p. 54)
• Scaling Requirements (p. 54)
• Access Requirements (p. 54)
• Region and Availability Zone Requirements (p. 55)

Memory and Processor Requirements
The basic building block of Amazon ElastiCache is the node. Nodes are configured singularly or in groupings to form clusters. When determining the node type to use for your cluster, take the cluster's node configuration and the amount of data you have to store into consideration.

The Memcached engine is multi-threaded, so a node's number of cores impacts the compute power available to the cluster.

Memcached Cluster Configuration
ElastiCache for Memcached clusters are comprised of from 1 to 20 nodes. The data in a Memcached cluster is partitioned across the nodes in the cluster. Your application connects with a Memcached cluster using a network address called an Endpoint. Each node in a Memcached cluster has its own endpoint which your application uses to read from or write to the specific node. In addition to the node endpoints, the Memcached cluster itself has an endpoint called the configuration endpoint. Your application can use this endpoint to read from or write to the cluster, leaving the determination of which node to read from or write to up to Automatically Identify Nodes in your Memcached Cluster (p. 55).

For more information, see Managing Your ElastiCache Clusters (p. 51).

Scaling Requirements
All clusters can be scaled up by creating a new cluster with the new, larger node type. When scaling up a Memcached cluster the new cluster will start out empty.

Amazon ElastiCache for Memcached clusters can be scaled out or in. To scale a Memcached cluster out or in you merely add or remove nodes from the cluster. If you have enabled Automatic Discovery and your application is connecting to the cluster's configuration endpoint, you do not need to make any changes in your application when you add or remove nodes.

For more information, see Scaling ElastiCache for Memcached Clusters (p. 132) in this guide.

Access Requirements
By design, Amazon ElastiCache clusters are accessed from Amazon EC2 instances. Network access to an ElastiCache cluster is limited to the user account that created the cluster. Therefore, before you can
access a cluster from an Amazon EC2 instance, you must authorize the Amazon EC2 instance to access the cluster. The steps to do this vary, depending upon whether you launched into EC2-VPC or EC2-Classic.

If you launched your cluster into EC2-VPC you need to grant network ingress to the cluster. If you launched your cluster into EC2-Classic you need to grant the Amazon Elastic Compute Cloud security group associated with the instance access to your ElastiCache security group. For detailed instructions, see Step 2: Authorize Access (p. 19) in this guide.

Region and Availability Zone Requirements

Amazon ElastiCache supports all AWS regions. By locating your ElastiCache clusters in a region close to your application you can reduce latency. If your cluster has multiple nodes, locating your nodes in different Availability Zones can reduce the impact of failures on your cluster.

For more information, see the following:

- Choosing Regions and Availability Zones (p. 37)
- Mitigating Failures (p. 251)

Automatically Identify Nodes in your Memcached Cluster

For clusters running the Memcached engine, ElastiCache supports Auto Discovery—the ability for client programs to automatically identify all of the nodes in a cache cluster, and to initiate and maintain connections to all of these nodes.

Note
Auto Discovery is added for cache clusters running on Amazon ElastiCache Memcached.

With Auto Discovery, your application does not need to manually connect to individual cache nodes; instead, your application connects to one Memcached node and retrieves the list of nodes. From that list your application is aware of the rest of the nodes in the cluster and can connect to any of them. You do not need to hard code the individual cache node endpoints in your application.

All of the cache nodes in the cluster maintain a list of metadata about all of the other nodes. This metadata is updated whenever nodes are added or removed from the cluster.

Topics

- Benefits of Auto Discovery (p. 56)
- How Auto Discovery Works (p. 57)
- Using Auto Discovery (p. 60)
- Connecting to Cache Nodes Manually (p. 65)
- Adding Auto Discovery To Your Client Library (p. 66)
- ElastiCache Clients with Auto Discovery (p. 67)
Benefits of Auto Discovery

Auto Discovery offers the following benefits:

- When you increase the number of nodes in a cache cluster, the new nodes register themselves with the configuration endpoint and with all of the other nodes. When you remove nodes from the cache cluster, the departing nodes deregister themselves. In both cases, all of the other nodes in the cluster are updated with the latest cache node metadata.
- Cache node failures are automatically detected; failed nodes are automatically replaced.

  **Note**
  Until node replacement completes, the node will continue to fail.

- A client program only needs to connect to the configuration endpoint. After that, the Auto Discovery library connects to all of the other nodes in the cluster.
- Client programs poll the cluster once per minute (this interval can be adjusted if necessary). If there are any changes to the cluster configuration, such as new or deleted nodes, the client receives an updated list of metadata. Then the client connects to, or disconnects from, these nodes as needed.

Auto Discovery is enabled on all ElastiCache Memcached cache clusters. You do not need to reboot any of your cache nodes to use this feature.
How Auto Discovery Works

Topics
- Connecting to Cache Nodes (p. 57)
- Normal Cluster Operations (p. 58)
- Other Operations (p. 58)

This section describes how client applications use ElastiCache Cluster Client to manage cache node connections, and interact with data items in the cache.

Connecting to Cache Nodes

From the application's point of view, connecting to the cluster configuration endpoint is no different from connecting directly to an individual cache node. The following sequence diagram shows the process of connecting to cache nodes.

Process of Connecting to Cache Nodes

1. The application resolves the configuration endpoint’s DNS name. Because the configuration endpoint maintains CNAME entries for all of the cache nodes, the DNS name resolves to one of the nodes; the client can then connect to that node.

2. The client requests the configuration information for all of the other nodes. Since each node maintains configuration information for all of the nodes in the cluster, any node can pass configuration information to the client upon request.

3. The client receives the current list of cache node hostnames and IP addresses. It can then connect to all of the other nodes in the cluster.
Note
The client program refreshes its list of cache node hostnames and IP addresses once per minute. This polling interval can be adjusted if necessary.

Normal Cluster Operations

When the application has connected to all of the cache nodes, ElastiCache Cluster Client determines which nodes should store individual data items, and which nodes should be queried for those data items later. The following sequence diagram shows the process of normal cluster operations.

Process of Normal Cluster Operations

1. The application issues a get request for a particular data item, identified by its key.
2. The client uses a hashing algorithm against the key to determine which cache node contains the data item.
3. The data item is requested from the appropriate node.
4. The data item is returned to the application.

Other Operations

In some situations, you might make a change to a cluster's nodes. For example, you might add an additional node to accommodate additional demand, or delete a node to save money during periods of reduced demand. Or you might replace a node due to a node failure of one sort or another.
Creating a Cluster

When there is a change in the cluster that requires a metadata update to the cluster's endpoints, that change is made to all nodes at the same time. Thus the metadata in any given node is consistent with the metadata in all of the other nodes in the cluster.

In each of these cases, the metadata is consistent among all the nodes at all times since the metadata is updated at the same time for all nodes in the cluster. You should always use the configuration endpoint to obtain the endpoints of the various nodes in the cluster. By using the configuration endpoint, you ensure that you will not be obtaining endpoint data from a node that “disappears” on you.

**Adding a Node**

During the time that the node is being spun up, its endpoint is not included in the metadata. As soon as the node is available, it is added to the metadata of each of the cluster’s nodes. In this scenario, the metadata is consistent among all the nodes and you will be able to interact with the new node only after it is available. Before the node being available, you will not know about it and will interact with the nodes in your cluster the same as though the new node does not exist.

**Deleting a Node**

When a node is removed, its endpoint is first removed from the metadata and then the node is removed from the cluster. In this scenario the metadata in all the nodes is consistent and there is no time in which it will contain the endpoint for the node to be removed while the node is not available. During the node removal time it is not reported in the metadata and so your application will only be interacting with the n-1 remaining nodes, as though the node does not exist.

**Replacing a Node**

If a node fails, ElastiCache takes down that node and spins up a replacement. The replacement process takes a few minutes. During this time the metadata in all the nodes still shows the endpoint for the failed node, but any attempt to interact with the node will fail. Therefore, your logic should always include retry logic.
Using Auto Discovery

To begin using Auto Discovery, follow these steps:

- Step 1: Obtain the Configuration Endpoint (p. 60)
- Step 2: Download the ElastiCache Cluster Client (p. 61)
- Step 3: Modify Your Application Program (p. 61)

Step 1: Obtain the Configuration Endpoint

To connect to a cluster, client programs must know the cluster configuration endpoint. See the topic Finding a Cluster's Endpoints (Console) (p. 128)

You can also use the `aws elasticache describe-cache-clusters` command with the `--show-cache-node-info` parameter:

Whatever method you use to find the cluster's endpoints, the configuration endpoint will always have `.cfg` in its address.

Example Finding endpoints using the AWS CLI for ElastiCache

For Linux, macOS, or Unix:

```bash
aws elasticache describe-cache-clusters \
    --cache-cluster-id mycluster \
    --show-cache-node-info
```

For Windows:

```bash
aws elasticache describe-cache-clusters ^ \
    --cache-cluster-id mycluster ^ \
    --show-cache-node-info
```

This operation produces output similar to the following (JSON format):

```json
{
    "CacheClusters": [
        {
            "Engine": "memcached",
            "CacheNodes": [
                {
                    "CacheNodeId": "0001",
                    "Endpoint": {
                        "Port": 11211,
                        "Address": "mycluster.fnjyzo.cfg.0001.use1.cache.amazonaws.com"
                    },
                    "CacheNodeStatus": "available",
                    "ParameterGroupStatus": "in-sync",
                    "CustomerAvailabilityZone": "us-east-1e"
                },
                {
                    "CacheNodeId": "0002",
                    "Endpoint": {
                        "Port": 11211,
                        "Address": "mycluster.fnjyzo.cfg.0002.use1.cache.amazonaws.com"
                    },
                    "CacheNodeStatus": "available",
```

API Version 2015-02-02

60
Step 2: Download the ElastiCache Cluster Client

To take advantage of Auto Discovery, client programs must use the ElastiCache Cluster Client. The ElastiCache Cluster Client is available for Java, PHP, and .NET and contains all of the necessary logic for discovering and connecting to all of your cache nodes.

To download the ElastiCache Cluster Client

2. From the ElastiCache console, choose ElastiCache Cluster Client then choose Download.

The source code for the ElastiCache Cluster Client for Java is available at https://github.com/amazonwebservices/aws-elasticache-cluster-client-memcached-for-java. This library is based on the popular Spymemcached client. The ElastiCache Cluster Client is released under the Amazon Software License https://aws.amazon.com/asl. You are free to modify the source code as you see fit. You can even incorporate the code into other open source Memcached libraries, or into your own client code.

Note

To use the ElastiCache Cluster Client for PHP, you will first need to install it on your Amazon EC2 instance. For more information, see Installing the ElastiCache Cluster Client for PHP (p. 70).
To use the ElastiCache Cluster Client for .NET, you will first need to install it on your Amazon EC2 instance. For more information, see Installing the ElastiCache Cluster Client for .NET (p. 68).

Step 3: Modify Your Application Program

Modify your application program so that it uses Auto Discovery. The following sections show how to use the ElastiCache Cluster Client for Java, PHP, and .NET.
Important
When specifying the cluster's configuration endpoint, be sure that the endpoint has ".cfg" in its address as shown here. Do not use a CNAME or an endpoint without ".cfg" in it.

"mycluster.fnjyzo.cfg.use1.cache.amazonaws.com;"

Failure to explicitly specify the cluster's configuration endpoint results in configuring to a specific node.

Topics
- Using the ElastiCache Cluster Client for Java (p. 62)
- Using the ElastiCache Cluster Client for PHP (p. 62)
- Using the ElastiCache Cluster Client for .NET (p. 63)

Using the ElastiCache Cluster Client for Java

The program below demonstrates how to use the ElastiCache Cluster Client to connect to a cluster configuration endpoint and add a data item to the cache. Using Auto Discovery, the program connects to all of the nodes in the cluster without any further intervention.

```java
package com.amazon.elasticache;

import java.io.IOException;
import java.net.InetSocketAddress;

// Import the AWS-provided library with Auto Discovery support
import net.spy.memcached.MemcachedClient;

public class AutoDiscoveryDemo {
    public static void main(String[] args) throws IOException {
        String configEndpoint = "mycluster.fnjyzo.cfg.use1.cache.amazonaws.com;";
        Integer clusterPort = 11211;

        MemcachedClient client = new MemcachedClient(
            new InetSocketAddress(configEndpoint, clusterPort));
        // The client will connect to the other cache nodes automatically.

        // Store a data item for an hour.
        // The client will decide which cache host will store this item.
        client.set("theKey", 3600, "This is the data value");
    }
}
```

Using the ElastiCache Cluster Client for PHP

The program below demonstrates how to use the ElastiCache Cluster Client to connect to a cluster configuration endpoint and add a data item to the cache. Using Auto Discovery, the program will connect to all of the nodes in the cluster without any further intervention.

To use the ElastiCache Cluster Client for PHP, you will first need to install it on your Amazon EC2 instance. For more information, see Installing the ElastiCache Cluster Client for PHP (p. 70)

```php
<?php

/**
*/
```
* Sample PHP code to show how to integrate with the Amazon ElastiCache
* Auto Discovery feature.
*/

/* Configuration endpoint to use to initialize memcached client.
* This is only an example. */
$server_endpoint = "mycluster.fnjyo.cfg.use1.cache.amazonaws.com";

/* Port for connecting to the ElastiCache cluster.
* This is only an example */
$server_port = 11211;

/**
* The following will initialize a Memcached client to utilize the Auto Discovery feature.
*
* By configuring the client with the Dynamic client mode with single endpoint, the
* client will periodically use the configuration endpoint to retrieve the current cache
* cluster configuration. This allows scaling the cache cluster up or down in number of
* nodes
* without requiring any changes to the PHP application.
*
* By default the Memcached instances are destroyed at the end of the request.
* To create an instance that persists between requests,
* use persistent_id to specify a unique ID for the instance.
* All instances created with the same persistent_id will share the same connection.
*/
$dynamic_client = new Memcached('persistent-id');
$dynamic_client->setOption(Memcached::OPT_CLIENT_MODE, Memcached::DYNAMIC_CLIENT_MODE);
$dynamic_client->addServer($server_endpoint, $server_port);

/**
* Store the data for 60 seconds in the cluster.
* The client will decide which cache host will store this item.
*/
$dynamic_client->set('key', 'value', 60);

/**
* Configuring the client with Static client mode disables the usage of Auto Discovery
* and the client operates as it did before the introduction of Auto Discovery.
* The user can then add a list of server endpoints.
*/
$static_client = new Memcached('persistent-id');
$static_client->setOption(Memcached::OPT_CLIENT_MODE, Memcached::STATIC_CLIENT_MODE);
$static_client->addServer($server_endpoint, $server_port);

/**
* Store the data without expiration.
* The client will decide which cache host will store this item.
*/
$static_client->set('key', 'value');
?>

Using the ElastiCache Cluster Client for .NET


.NET applications typically get their configurations from their config file. The following is a sample application config file.

```xml
<?xml version="1.0" encoding="utf-8"?>
<configuration>
  <configSections>
```
The C# program below demonstrates how to use the ElastiCache Cluster Client to connect to a cluster configuration endpoint and add a data item to the cache. Using Auto Discovery, the program will connect to all of the nodes in the cluster without any further intervention.

```csharp
// *****************
// Sample C# code to show how to integrate with the Amazon ElastiCache Auto Discovery feature.
using System;
using Amazon.ElastiCacheCluster;
using Enyim.Caching;
using Enyim.Caching.Memcached;

class DotNetAutoDiscoveryDemo {

    public static void Main(string[] args) {

        // instantiate a new client.
        ElastiCacheClusterConfig config = new ElastiCacheClusterConfig();
        MemcachedClient memClient = new MemcachedClient(config);

        // Store the data for 3600 seconds (1 hour) in the cluster.
        // The client will decide which cache host will store this item.
        memClient.Store(StoreMode.Set, 3600, "This is the data value.");

    } // end Main

} // end class DotNetAutoDiscoveryDemo
```
Connecting to Cache Nodes Manually

If your client program does not use Auto Discovery, it can manually connect to each of the cache nodes. This is the default behavior for Memcached clients.

You can obtain a list of cache node hostnames and port numbers from the AWS Management Console. You can also use the AWS CLI `aws elasticache describe-cache-clusters` command with the `--show-cache-node-info` parameter.

Example

The following Java code snippet shows how to connect to all of the nodes in a four-node cache cluster:

```java
ArrayList<String> cacheNodes = new ArrayList<String>(
    Arrays.asList(
        "mycachecluster.fnjyzo.0001.use1.cache.amazonaws.com:11211",
        "mycachecluster.fnjyzo.0002.use1.cache.amazonaws.com:11211",
        "mycachecluster.fnjyzo.0003.use1.cache.amazonaws.com:11211",
        "mycachecluster.fnjyzo.0004.use1.cache.amazonaws.com:11211"));

MemcachedClient cache = new MemcachedClient(AddrUtil.getAddresses(cacheNodes));
```

Important

If you scale up or scale down your cache cluster by adding or removing nodes, you will need to update the list of nodes in the client code.
Adding Auto Discovery To Your Client Library

The configuration information for Auto Discovery is stored redundantly in each cache cluster node. Client applications can query any cache node and obtain the configuration information for all of the nodes in the cluster.

The way in which an application does this depends upon the cache engine version:

- If the cache engine version is **1.4.14 or higher**, use the `config` command.
- If the cache engine version is **lower than 1.4.14**, use the `get AmazonElastiCache:cluster` command.

The outputs from these two commands are identical, and are described in the [Output Format](p. 67) section below.

**Cache Engine Version 1.4.14 or Higher**

For cache engine version 1.4.14 or higher, use the `config` command. This command has been added to the Memcached ASCII and binary protocols by ElastiCache, and is implemented in the ElastiCache Cluster Client. If you want to use Auto Discovery with another client library, then that library will need to be extended to support the `config` command.

**Note**
The following documentation pertains to the ASCII protocol; however, the `config` command supports both ASCII and binary. If you want to add Auto Discovery support using the binary protocol, refer to the [source code for the ElastiCache Cluster Client](#).

**Syntax**

```
config [sub-command] [key]
```

**Options**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub-command</td>
<td>The sub-command used to interact with a cache node. For Auto Discovery, this sub-command is <code>get</code>.</td>
<td>Yes</td>
</tr>
<tr>
<td>key</td>
<td>The key under which the cluster configuration is stored. For Auto Discovery, this key is named <code>cluster</code>.</td>
<td>Yes</td>
</tr>
</tbody>
</table>

To get the cluster configuration information, use the following command:

```
config get cluster
```

**Cache Engine Version Lower Than 1.4.14**

To get the cluster configuration information, use the following command:

```
get AmazonElastiCache:cluster
```

**Note**
Do not tamper with the "AmazonElastiCache:cluster" key, since this is where the cluster configuration information resides. If you do overwrite this key, then the client may be incorrectly configured for a brief period of time (no more than 15 seconds) before ElastiCache automatically and correctly updates the configuration information.
Output Format

Whether you use `config get cluster` or `get AmazonElastiCache:cluster`, the reply consists of two lines:

- The version number of the configuration information. Each time a node is added or removed from the cache cluster, the version number increases by one.
- A list of cache nodes. Each node in the list is represented by a `hostname|ip-address|port` group, and each node is delimited by a space.

A carriage return and a linefeed character (CR + LF) appears at the end of each line. The data line contains a linefeed character (LF) at the end, to which the CR + LF is added. The config version line is terminated by LF without the CR.

A cache cluster containing three nodes would be represented as follows:

```
configversion
hostname|ip-address|port hostname|ip-address|port hostname|ip-address|port

```

Each node is shown with both the CNAME and the private IP address. The CNAME will always be present; if the private IP address is not available, it will not be shown; however, the pipe characters "|" will still be printed.

Example

Here is an example of the payload returned when you query the configuration information:

```
CONFIG cluster 0 147
12
myCluster.pc4ldq.0001.use1.cache.amazonaws.com|10.82.235.120|11211
myCluster.pc4ldq.0002.use1.cache.amazonaws.com|10.80.249.27|11211
END
```

Note

- The second line indicates that the configuration information has been modified twelve times so far.
- In the third line, the list of nodes is in alphabetical order by hostname. This ordering might be in a different sequence from what you are currently using in your client application.

ElastiCache Clients with Auto Discovery

This section discusses installing and configuring the ElastiCache PHP and .NET clients.

Topics

- Installing & Compiling Cluster Clients (p. 67)
- Configuring ElastiCache Clients (p. 79)

Installing & Compiling Cluster Clients

This section covers installing, configuring, and compiling the PHP and .NET Amazon ElastiCache auto discovery cluster clients.

Topics
Creating a Cluster

- Installing the ElastiCache Cluster Client for .NET (p. 68)
- Installing the ElastiCache Cluster Client for PHP (p. 70)
- Compiling the Source Code for the ElastiCache Cluster Client for PHP (p. 77)

Installing the ElastiCache Cluster Client for .NET

You can find the ElastiCache .NET Cluster Client code as open source at https://github.com/awslabs/elasticache-cluster-config-net.

This section describes how to install, update, and remove the .NET components for the ElastiCache Cluster Client on Amazon EC2 instances. For more information about auto discovery, see Automatically Identify Nodes in your Memcached Cluster (p. 55). For sample .NET code to use the client, see Using the ElastiCache Cluster Client for .NET (p. 63).

Topics
- Installing .NET (p. 68)
- Download the ElastiCache .NET Cluster Client for ElastiCache (p. 68)
- Install AWS Assemblies with NuGet (p. 68)

Installing .NET

You must have .NET 3.5 or later installed to use the AWS .NET SDK for ElastiCache. If you don't have .NET 3.5 or later, you can download and install the latest version from http://www.microsoft.com/net.

Download the ElastiCache .NET Cluster Client for ElastiCache

To download the ElastiCache .NET cluster client

2. On the navigation pane, click ElastiCache Cluster Client.  
3. In the Download ElastiCache Memcached Cluster Client list, select .NET, and then click Download.

Install AWS Assemblies with NuGet

NuGet is a package management system for the .NET platform. NuGet is aware of assembly dependencies and installs all required files automatically. NuGet installed assemblies are stored with your solution, rather than in a central location such as Program Files, so you can install versions specific to an application without creating compatibility issues.

Installing NuGet

NuGet can be installed from the Installation Gallery on MSDN; see https://visualstudiogallery.msdn.microsoft.com/27077b70-9dad-4c64-adcf-c7cf6bc9970c. If you are using Visual Studio 2010 or later, NuGet is automatically installed.

You can use NuGet from either Solution Explorer or Package Manager Console.

Using NuGet from Solution Explorer

To use NuGet from Solution Explorer in Visual Studio 2010

1. From the Tools menu, select Library Package Manager.  
2. Click Package Manager Console.
To use NuGet from Solution Explorer in Visual Studio 2012 or Visual Studio 2013

1. From the Tools menu, select NuGet Package Manager.
2. Click Package Manager Console.

From the command line, you can install the assemblies using Install-Package, as shown following.

```
Install-Package Amazon.ElastiCacheCluster
```

To see a page for every package that is available through NuGet, such as the AWSSDK and AWS.Extensions assemblies, see the NuGet website at http://www.nuget.org. The page for each package includes a sample command line for installing the package using the console and a list of the previous versions of the package that are available through NuGet.

For more information on Package Manager Console commands, see http://nuget.codeplex.com/wikipage?title=Package%20Manager%20Console%20Command%20Reference%20%28v1.3%29.
Installing the ElastiCache Cluster Client for PHP

This section describes how to install, update, and remove the PHP components for the ElastiCache Cluster Client on Amazon EC2 instances. For more information about Auto Discovery, see Automatically Identify Nodes in your Memcached Cluster (p. 55). For sample PHP code to use the client, see Using the ElastiCache Cluster Client for PHP (p. 62).

Topics
- Downloading the Installation Package (p. 70)
- For Users Who Already Have php-memcached Extension Installed (p. 71)
- Installation Steps for New Users (p. 71)
- Removing the PHP Cluster Client (p. 76)

Downloading the Installation Package

To ensure that you use the correct version of the ElastiCache Cluster Client for PHP, you will need to know what version of PHP is installed on your Amazon EC2 instance. You will also need to know whether your Amazon EC2 instance is running a 64-bit or 32-bit version of Linux.

To determine the PHP version installed on your Amazon EC2 instance

- At the command prompt, run the following command:

  ```bash
  php -v
  ```

  The PHP version will be shown in the output, as in this example:

  ```
  PHP 5.4.10 (cli) (built: Jan 11 2013 14:48:57)
  Copyright (c) 1997-2012 The PHP Group
  Zend Engine v2.4.0, Copyright (c) 1998-2012 Zend Technologies
  ```

  **Note**
  If your PHP and Memcached versions are incompatible, you will get an error message something like the following:

  ```
  PHP Warning: PHP Startup: memcached: Unable to initialize module
  Module compiled with module API=20100525
  PHP compiled with module API=20131226
  These options need to match
  in Unknown on line 0
  ```

  If this happens, you need to compile the module from the source code. For more information, see Compiling the Source Code for the ElastiCache Cluster Client for PHP (p. 77).

To determine your Amazon EC2 AMI architecture (64-bit or 32-bit)

1. Sign in to the AWS Management Console and open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
2. In the **Instances** list, click your Amazon EC2 instance.
3. In the **Description** tab, look for the **AMI** field. A 64-bit instance should have x86_64 as part of the description; for a 32-bit instance, look for i386 or i686 in this field.

You are now ready to download the ElastiCache Cluster Client.
To download the ElastiCache Cluster Client for PHP

2. From the ElastiCache console, choose ElastiCache Cluster Client.
3. From the Download ElastiCache Memcached Cluster Client list, choose the ElastiCache Cluster Client that matches your PHP version and AMI architecture, then choose the Download button.

For Users Who Already Have php-memcached Extension Installed

To update the php-memcached installation

1. Remove the previous installation of the Memcached extension for PHP as described by the topic Removing the PHP Cluster Client (p. 76).
2. Install the new ElastiCache php-memcached extension as described previously in Installation Steps for New Users (p. 71).

Installation Steps for New Users

Topics
- Installing PHP 7.x for New Users (p. 71)
- Installing PHP 5.x for New Users (p. 73)

Installing PHP 7.x for New Users

Topics
- To install PHP 7 on a Ubuntu Server 14.04 LTS AMI (64-bit and 32-bit) (p. 71)
- To install PHP 7 on an Amazon Linux 201609 AMI (p. 72)
- To install PHP 7 on an SUSE Linux AMI (p. 72)

To install PHP 7 on a Ubuntu Server 14.04 LTS AMI (64-bit and 32-bit)

1. Launch a new instance from the AMI.
2. Run the following commands:
   ```bash
   sudo apt-get update
   sudo apt-get install gcc g++
   ```
3. Install PHP 7.
   ```bash
   sudo yum install php70
   ```
4. Download the Amazon ElastiCache Cluster Client.
   ```bash
   wget https://elasticache-downloads.s3.amazonaws.com/ClusterClient/PHP-7.0/latest-64bit
   ```
5. Extract latest-64bit.
   ```bash
   tar -zxvf latest-64bit
   ```
To install PHP 7 on an Amazon Linux 201609 AMI

1. Launch a new instance from the AMI.
2. Run the following command:

   ```bash
   sudo yum install gcc-c++
   ```
3. Install PHP 7.

   ```bash
   sudo yum install php70
   ```
4. Download the Amazon ElastiCache Cluster Client.

   ```bash
   wget https://elasticache-downloads.s3.amazonaws.com/ClusterClient/PHP-7.0/latest-64bit
   ```
5. Extract latest-64bit.

   ```bash
   tar -zxvf latest-64bit
   ```
6. With root permission, copy the extracted artifact file `amazon-elasticache-cluster-client.so` into `/usr/lib64/php/7.0/modules/`.

   ```bash
   sudo mv artifact/amazon-elasticache-cluster-client.so /usr/lib64/php/7.0/modules/
   ```
7. Create the `50-memcached.ini` file.

   ```bash
   echo "extension=amazon-elasticache-cluster-client.so" | sudo tee --append /etc/php-7.0.d/50-memcached.ini
   ```
8. Start or restart your Apache server.

   ```bash
   sudo /etc/init.d/httpd start
   ```

To install PHP 7 on an SUSE Linux AMI

1. Launch a new instance from the AMI.
2. Run the following command:
3. Install PHP 7.

```
sudo yum install php70
```

4. Download the Amazon ElastiCache Cluster Client.

```
wget https://elasticache-downloads.s3.amazonaws.com/ClusterClient/PHP-7.0/latest-64bit
```

5. Extract latest-64bit.

```
tar -zxvf latest-64bit
```

6. With root permission, copy the extracted artifact file amazon-elasticache-cluster-client.so into /usr/lib64/php7/extensions/.

```
sudo mv artifact/amazon-elasticache-cluster-client.so /usr/lib64/php7/extensions/
```

7. Insert the line `extension=amazon-elasticache-cluster-client.so` into the file /etc/php7/cli/php.ini.

```
echo "extension=amazon-elasticache-cluster-client.so" | sudo tee --append /etc/php7/cli/php.ini
```

8. Start or restart your Apache server.

```
sudo /etc/init.d/httpd start
```

---

**Installing PHP 5.x for New Users**

**Topics**

- To install PHP 5 on an Amazon Linux AMI 2014.03 (64-bit and 32-bit) (p. 73)
- To install PHP 5 on a Red Hat Enterprise Linux 7.0 AMI (64-bit and 32-bit) (p. 74)
- To install PHP 5 on a Ubuntu Server 14.04 LTS AMI (64-bit and 32-bit) (p. 74)
- To install PHP 5 for SUSE Linux Enterprise Server 11 AMI (64-bit or 32-bit) (p. 75)
- Other Linux distributions (p. 76)

**To install PHP 5 on an Amazon Linux AMI 2014.03 (64-bit and 32-bit)**

1. Launch an Amazon Linux instance (either 64-bit or 32-bit) and log into it.
2. Install PHP dependencies:

```
$ sudo yum install gcc-c++ php php-pear
```

3. Download the correct php-memcached package for your Amazon EC2 instance and PHP version. For more information, see Downloading the Installation Package (p. 70).

4. Install php-memcached. The URI should be the download path for the installation package:

```
$ sudo pecl install <package download path>
```
Here is a sample installation command for PHP 5.4, 64-bit Linux. In this sample, replace $X.Y.Z$ with the actual version number:

```
$ sudo pecl install /home/AmazonElastiCacheClusterClient-X.Y.Z-PHP54-64bit.tgz
```

**Note**
Be sure to use the latest version of the install artifact.

5. With root/sudo permission, add a new file named `memcached.ini` in the `/etc/php.d` directory, and insert "extension=amazon-elasticache-cluster-client.so" in the file:

```
$ echo "extension=amazon-elasticache-cluster-client.so" | sudo tee --append /etc/php.d/memcached.ini
```

6. Start or restart your Apache server.

```
sudo /etc/init.d/httpd start
```

---

**To install PHP 5 on a Red Hat Enterprise Linux 7.0 AMI (64-bit and 32-bit)**

1. Launch a Red Hat Enterprise Linux instance (either 64-bit or 32-bit) and log into it.
2. Install PHP dependencies:

```
sudo yum install gcc-c++ php php-pear
```

3. Download the correct `php-memcached` package for your Amazon EC2 instance and PHP version. For more information, see Downloading the Installation Package (p. 70).
4. Install `php-memcached`. The URI should be the download path for the installation package:

```
sudo pecl install <package download path>
```


```
echo "extension=amazon-elasticache-cluster-client.so" | sudo tee --append /etc/php.d/memcached.ini
```

6. Start or restart your Apache server.

```
sudo /etc/init.d/httpd start
```

---

**To install PHP 5 on a Ubuntu Server 14.04 LTS AMI (64-bit and 32-bit)**

1. Launch an Ubuntu Linux instance (either 64-bit or 32-bit) and log into it.
2. Install PHP dependencies:

```
sudo apt-get update
sudo apt-get install gcc g++ php5 php-pear
```

API Version 2015-02-02
3. Download the correct `php-memcached` package for your Amazon EC2 instance and PHP version. For more information, see Downloading the Installation Package (p. 70).

4. Install `php-memcached`. The URI should be the download path for the installation package.

```
$ sudo pecl install <package download path>
```

**Note**

This installation step installs the build artifact `amazon-elasticache-cluster-client.so` into the `/usr/lib/php5/20121212*` directory. Verify the absolute path of the build artifact, because you need it in the next step.

If the previous command doesn't work, you need to manually extract the PHP client artifact `amazon-elasticache-cluster-client.so` from the downloaded `*.tgz` file, and copy it to the `/usr/lib/php5/20121212*` directory.

```
$ tar -xvf <package download path>
cp amazon-elasticache-cluster-client.so /usr/lib/php5/20121212/
```


```
$ echo "extension=<absolute path to amazon-elasticache-cluster-client.so>" | sudo tee --append /etc/php5/cli/conf.d/memcached.ini
```

6. Start or restart your Apache server.

```
sudo /etc/init.d/httpd start
```

---

**To install PHP 5 for SUSE Linux Enterprise Server 11 AMI (64-bit or 32-bit)**

1. Launch a SUSE Linux instance (either 64-bit or 32-bit) and log into it.

2. Install PHP dependencies:

```
$ sudo zypper install gcc php53-devel
```

3. Download the correct `php-memcached` package for your Amazon EC2 instance and PHP version. For more information, see Downloading the Installation Package (p. 70).

4. Install `php-memcached`. The URI should be the download path for the installation package.

```
$ sudo pecl install <package download path>
```


```
$ echo "extension=amazon-elasticache-cluster-client.so" | sudo tee --append /etc/php5/conf.d/memcached.ini
```

6. Start or restart your Apache server.

```
sudo /etc/init.d/httpd start
```

---

API Version 2015-02-02

75
**Note**

If Step 5 doesn't work for any of the previous platforms, verify the install path for `amazon-elasticache-cluster-client.so`. Also, specify the full path of the binary in the extension. In addition, verify that the PHP in use is a supported version. We support versions 5.3 through 5.5.

**Other Linux distributions**

On some systems, notably CentOS7 and Red Hat Enterprise Linux (RHEL) 7.1, `libsasl2.so.3` has replaced `libsasl2.so.2`. On those systems, when you load the ElastiCache cluster client, it attempts and fails to find and load `libsasl2.so.2`. To resolve this issue, create a symbolic link to `libsasl2.so.3` so that when the client attempts to load `libsasl2.so.2`, it is redirected to `libsasl2.so.3`. The following code creates this symbolic link.

```
cd /usr/lib64
$ sudo ln -s libsasl2.so.3 libsasl2.so.2
```

**Removing the PHP Cluster Client**

**Topics**

- Removing an earlier version of PHP 7 (p. 76)
- Removing an earlier version of PHP 5 (p. 76)

**Removing an earlier version of PHP 7**

1. Remove the `amazon-elasticache-cluster-client.so` file from the appropriate PHP lib directory as previously indicated in the installation instructions. See the section for your installation at For Users Who Already Have `php-memcached` Extension Installed (p. 71).

2. Remove the line `extension=amazon-elasticache-cluster-client.so` from the `php.ini` file.

3. Start or restart your Apache server.

```
sudo /etc/init.d/httpd start
```

**Removing an earlier version of PHP 5**

1. Remove the `php-memcached` extension:

```
sudo pecl uninstall __uri/AmazonElastiCacheClusterClient
```

2. Remove the `memcached.ini` file added in the appropriate directory as indicated in the previous installation steps.
Compiling the Source Code for the ElastiCache Cluster Client for PHP

This section covers how to obtain and compile the source code for the ElastiCache Cluster Client for PHP.

There are two packages you need to pull from GitHub and compile; aws-elasticache-cluster-client-libmemcached and aws-elasticache-cluster-client-memcached-for-php.

Topics
- Compiling the libmemcached Library (p. 77)
- Compiling the ElastiCache Memcached Auto Discovery Client for PHP (p. 77)

Compiling the libmemcached Library

To compile the aws-elasticache-cluster-client-libmemcached library
1. Launch an Amazon EC2 instance.
2. Install the library dependencies.
   - On Amazon Linux 201509 AMI
     ```
     sudo yum install gcc gcc-c++ autoconf libevent-devel
     ```
   - On Ubuntu 14.04 AMI
     ```
     sudo apt-get update
     sudo apt-get install libevent-dev gcc g++ make autoconf libsasl2-dev
     ```
3. Pull the repository and compile the code.
   ```
   Download and expand https://github.com/awslabs/aws-elasticache-cluster-client-libmemcached/archive/v1.0.18.tar.gz
   cd aws-elasticache-cluster-client-libmemcached
   mkdir BUILD
   cd BUILD
   ../configure --prefix=<libmemcached-install-directory> --with-pic
   make
   sudo make install
   ```

Compiling the ElastiCache Memcached Auto Discovery Client for PHP

The following sections describe how to compile the ElastiCache Memcached Auto Discovery Client

Topics
- Compiling the ElastiCache Memcached Client for PHP 7 (p. 77)
- Compiling the ElastiCache Memcached Client for PHP 5 (p. 78)

Compiling the ElastiCache Memcached Client for PHP 7

Run the following set of commands under the code directory.
```
git clone https://github.com/awslabs/aws-elasticache-cluster-client-memcached-for-php.git
cd aws-elasticache-cluster-client-memcached-for-php
git checkout php7
sudo yum install php70-devel
phpize
```
Creating a Cluster

./configure --with-libmemcached-dir=<libmemcached-install-directory> --disable-memcached-sasl
make
make install

Note
You can statically link the libmemcached library into the PHP binary so it can be ported across various Linux platforms. To do that, run the following command before make:

```bash
sed -i "s#-lmemcached#<libmemcached-install-directory>/lib/libmemcached.a -lcrypt -lpthread -lm -lstdc++ -lsasl2#" Makefile
```

Compiling the ElastiCache Memcached Client for PHP 5

Compile the `aws-elasticache-cluster-client-memcached-for-php` by running the following commands under the `aws-elasticache-cluster-client-memcached-for-php` folder.

```bash
git clone https://github.com/awslabs/aws-elasticache-cluster-client-memcached-for-php.git
cd aws-elasticache-cluster-client-memcached-for-php
sudo yum install zlib-devel phpize
./configure --with-libmemcached-dir=<libmemcached-install-directory>
make
make install
```
Configuring ElastiCache Clients

An ElastiCache cluster is protocol-compliant with Memcached. The code, applications, and most popular tools that you use today with your existing Memcached environment will work seamlessly with the service.

This section discusses specific considerations for connecting to cache nodes in ElastiCache.

Topics
- Finding Node Endpoints and Port Numbers (p. 79)
- Connecting for Using Auto Discovery (p. 80)
- DNS Names and Underlying IP (p. 80)

Finding Node Endpoints and Port Numbers

To connect to a cache node, your application needs to know the endpoint and port number for that node.

Finding Node Endpoints and Port Numbers (Console)

To determine node endpoints and port numbers
1. Sign in to the Amazon ElastiCache Management Console and choose the engine running on your cluster.
   A list of all clusters running the chosen engine appears.
2. Continue below for the engine and configuration you're running.
3. Choose the name of the cluster of interest.
4. Locate the Port and Endpoint columns for the node you're interested in.

Finding Node Endpoints and Port Numbers (AWS CLI)

To determine cache node endpoints and port numbers, use the command `describe-cache-clusters` with the `--show-cache-node-info` parameter.

```bash
aws elasticache describe-cache-clusters --show-cache-node-info
```

The fully qualified DNS names and port numbers are in the Endpoint section of the output.

Finding Cache Node Endpoints and Port Numbers (ElastiCache API)

To determine cache node endpoints and port numbers, use the action `DescribeCacheClusters` with the `ShowCacheNodeInfo=true` parameter.

Example

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheClusters
&ShowCacheNodeInfo=true
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20140421T220302Z
&Version=2014-09-30
&X-Amz-Algorithm=AWS4-HMAC-SHA256
&X-Amz-Credential=<credential>
&X-Amz-Date=20140421T220302Z
```

API Version 2015-02-02
Connecting for Using Auto Discovery

If your applications use Auto Discovery, you only need to know the configuration endpoint for the cluster, rather than the individual endpoints for each cache node. For more information, see Automatically Identify Nodes in your Memcached Cluster (p. 55).

Note
At this time, Auto Discovery is only available for cache clusters running Memcached.

DNS Names and Underlying IP

Clients maintain a server list containing the addresses and ports of the servers holding the cache data. When using ElastiCache, the DescribeCacheClusters API (or the describe-cache-clusters command line utility) returns a fully qualified DNS entry and port number that can be used for the server list.

Important
It is important that client applications are configured to frequently resolve DNS names of cache nodes when they attempt to connect to a cache node endpoint.

VPC Installations

ElastiCache ensures that both the DNS name and the IP address of the cache node remain the same when cache nodes are recovered in case of failure.

Non-VPC Installations

ElastiCache ensures that the DNS name of a cache node is unchanged when cache nodes are recovered in case of failure; however, the underlying IP address of the cache node can change.

Most client libraries support persistent cache node connections by default. We recommend using persistent cache node connections when using ElastiCache. Client-side DNS caching can occur in multiple places, including client libraries, the language runtime, or the client operating system. You should review your application configuration at each layer to ensure that you are frequently resolving IP addresses for your cache nodes.

Choosing Your Node Size

The node size you select for your cluster impacts costs, performance, and fault tolerance.

Choosing Your Memcached Node Size

Memcached clusters contain one or more nodes with the cluster's data partitioned across the nodes. Because of this, the memory needs of the cluster and the memory of a node are related, but not the same. You can attain your needed cluster memory capacity by having a few large nodes or several smaller nodes. Further, as your needs change, you can add nodes to or remove nodes from the cluster and thus pay only for what you need.

The total memory capacity of your cluster is calculated by multiplying the number of nodes in the cluster by the RAM capacity of each node after deducting system overhead. The capacity of each node is based on the node type.

\[
\text{cluster_capacity} = \text{number_of_nodes} \times (\text{node_capacity} - \text{system_overhead})
\]

The number of nodes in the cluster is a key factor in the availability of your cluster running Memcached. The failure of a single node can have an impact on the availability of your application and the load on
your backend database. In such a case, ElastiCache provisions a replacement for a failed node and it gets repopulated. To reduce this availability impact, spread your memory and compute capacity over more nodes with smaller capacity, rather than using fewer high-capacity nodes.

In a scenario where you want to have 35 GB of cache memory, you can set up any of the following configurations:

- 11 cache.t2.medium nodes with 3.22 GB of memory and 2 threads each = 35.42 GB and 22 threads.
- 6 cache.m4.large nodes with 6.42 GB of memory and 2 threads each = 38.52 GB and 12 threads.
- 3 cache.r4.large nodes with 12.3 GB of memory and 2 threads each = 36.90 GB and 6 threads.
- 3 cache.m4.xlarge nodes with 14.28 GB of memory and 4 threads each = 42.84 GB and 12 threads.

### Comparing node options

<table>
<thead>
<tr>
<th>Node type</th>
<th>Memory</th>
<th>Cores</th>
<th>Hourly Cost *</th>
<th>Nodes Needed</th>
<th>Total Memory</th>
<th>Total Cores</th>
<th>Monthly Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>cache.t2.medium</td>
<td>3.22 GB</td>
<td>2</td>
<td>$0.068</td>
<td>11</td>
<td>35.42 GB</td>
<td>22</td>
<td>$538.56</td>
</tr>
<tr>
<td>cache.m4.large</td>
<td>6.42 GB</td>
<td>2</td>
<td>$0.156</td>
<td>6</td>
<td>38.52 GB</td>
<td>12</td>
<td>$673.92</td>
</tr>
<tr>
<td>cache.m4.xlarge</td>
<td>4.28 GB</td>
<td>4</td>
<td>$0.311</td>
<td>3</td>
<td>42.84 GB</td>
<td>12</td>
<td>$671.76</td>
</tr>
<tr>
<td>cache.m5.xlarge</td>
<td>2.93 GB</td>
<td>4</td>
<td>$0.311</td>
<td>3</td>
<td>38.81 GB</td>
<td>12</td>
<td>$671.76</td>
</tr>
<tr>
<td>cache.r4.large</td>
<td>12.3 GB</td>
<td>2</td>
<td>$0.228</td>
<td>3</td>
<td>36.9 GB</td>
<td>6</td>
<td>$492.48</td>
</tr>
<tr>
<td>cache.r5.large</td>
<td>13.07 GB</td>
<td>2</td>
<td>$0.216</td>
<td>3</td>
<td>39.22 GB</td>
<td>6</td>
<td>$466.56</td>
</tr>
</tbody>
</table>

* Hourly cost per node as of October 22, 2018.

Monthly cost at 100% usage for 30 days (720 hours).

These options each provide similar memory capacity but different computational capacity and cost. To compare the costs of your specific options, see Amazon ElastiCache Pricing.

For clusters running Memcached, some of the available memory on each node is used for connection overhead. For more information, see Memcached Connection Overhead (p. 158)

Using multiple nodes requires spreading the keys across them. Each node has its own endpoint. For easy endpoint management, you can use the ElastiCache the Auto Discovery feature, which enables client programs to automatically identify all of the nodes in a cluster. For more information, see Automatically Identify Nodes in your Memcached Cluster (p. 55).

In some cases, you might be unsure how much capacity you need. If so, for testing we recommend starting with one cache.m5.large node. Then monitor the memory usage, CPU utilization, and cache hit rate with the ElastiCache metrics that are published to Amazon CloudWatch. For more information on CloudWatch metrics for ElastiCache, see Monitoring Use with Metrics (p. 162). For production and larger workloads, the R5 nodes provide the best performance and RAM cost value.

If your cluster doesn’t have the hit rate that you want, you can easily add more nodes to increase the total available memory in your cluster.

If your cluster is bound by CPU but has sufficient hit rate, set up a new cluster with a node type that provides more compute power.
Creating a Memcached Cluster (Console)

When you use the Memcached engine, Amazon ElastiCache supports horizontally partitioning your data over multiple nodes. Memcached enables auto discovery so you don't need to keep track of the endpoints for each node. Memcached tracks each node's endpoint, updating the endpoint list as nodes are added and removed. All your application needs to interact with the cluster is the configuration endpoint. For more information on auto discovery, see Automatically Identify Nodes in your Memcached Cluster (p. 55).

To create a Memcached cluster using the ElastiCache console:

1. Sign in to the AWS Management Console and open the Amazon ElastiCache console at https://console.aws.amazon.com/elasticache/.
2. From the list in the upper-right corner, choose the AWS Region you want to launch this cluster in.
3. Choose Memcached from the navigation pane.
4. Choose Create.
5. For Cluster engine, choose Memcached. Choosing Memcached will create a Memcached cluster that looks something like this. The number of nodes is determined by the number of nodes you choose in Step 5.f (up to a maximum of 20).

6. Complete the Memcached settings section.
   a. In Name, type in a name for your cluster.

      Cluster naming constraints are as follows:
      • Must contain 1–40 alphanumeric characters or hyphens.
      • Must begin with a letter.
      • Can't contain two consecutive hyphens.
      • Can't end with a hyphen.

     b. For Engine version compatibility, choose the Memcached engine version you want this cluster to run. Unless you have a specific reason to run an older version, we recommend that you choose the latest version.

        Important
        You can upgrade to newer engine versions. For more information, see Upgrading Engine Versions (p. 36). Any change in Memcached engine versions is a disruptive process in which you lose your cluster data.
     c. In Port, accept the default port, 11211. If you have a reason to use a different port, type the port number.
   d. For Parameter group, choose the default parameter group, choose the parameter group you want to use with this cluster, or choose Create new to create a new parameter group to use with this cluster.

   Parameter groups control the run-time parameters of your cluster. For more information on parameter groups, see Memcached Specific Parameters (p. 151) and Creating a Parameter Group (p. 137).
e. For **Node type**, click the down arrow (▲). In the **Change node type** dialog box, choose the **Instance family** of the node type you want, choose the node type you want to use for this cluster, and then choose **Save**.

For more information, see Choosing Your Memcached Node Size (p. 80).

f. For **Number of nodes**, choose the number of nodes you want for this cluster. You will partition your data across the cluster's nodes.

If you need to change the number of nodes later, scaling horizontally is quite easy with Memcached. For more information, see Scaling ElastiCache for Memcached Clusters (p. 132)

7. Click **Advanced Memcached settings** and complete the section.

a. For **Subnet group**, choose the subnet you want to apply to this cluster.

For more information, see Subnets and Subnet Groups (p. 214).

b. For **Availability zone(s)**, you have two options:

- **No preference** – ElastiCache chooses the Availability Zone for each node in your cluster.
- **Specify availability zones** – Specify the Availability Zone for each node in your cluster.

If you chose to specify the Availability Zones, for each node choose an Availability Zone from the list to the right of each node name.

We recommend locating your nodes in multiple Availability Zones for improved fault tolerance. For more information, see Mitigating Availability Zone Failures (p. 252).

For more information, see Choosing Regions and Availability Zones (p. 37).

c. For **Security groups**, choose the security groups you want to apply to this cluster.

For more information, see Amazon VPCs and ElastiCache Security (p. 198).

d. The **Maintenance window** is the time, generally an hour in length, each week when ElastiCache schedules system maintenance for your cluster. You can allow ElastiCache choose the day and time for your maintenance window (**No preference**), or you can choose the day, time, and duration yourself (**Specify maintenance window**). If you choose **Specify maintenance window**, choose the **Start day**, **Start time**, and **Duration** (in hours) for your maintenance window. All times are UCT times.

For more information, see Managing Maintenance (p. 41).

e. For **Notifications**, choose an existing Amazon Simple Notification Service (Amazon SNS) topic, or choose manual ARN input and type in the topic Amazon Resource Name (ARN). Amazon SNS allows you to push notifications to Internet-connected smart devices. The default is to disable notifications. For more information, see https://aws.amazon.com/sns/.

8. Review all your entries and choices, then go back and make any needed corrections. When you're ready, choose **Create** to launch your cluster.

As soon as your cluster's status is **available**, you can grant Amazon EC2 access to it, connect to it, and begin using it. For more information, see Step 2: Authorize Access (p. 19) and Step 3: Connect to a Cluster's Node (p. 20).

**Important**

As soon as your cluster becomes available, you're billed for each hour or partial hour that the cluster is active, even if you're not actively using it. To stop incurring charges for this cluster, you must delete it. See Deleting a Cluster (p. 105).
Creating a Cluster (AWS CLI)

To create a cluster using the AWS CLI, use the create-cache-cluster command.

**Important**
As soon as your cluster becomes available, you're billed for each hour or partial hour that the cluster is active, even if you're not actively using it. To stop incurring charges for this cluster, you must delete it. SeeDeleting a Cluster (p. 105).

**Topics**
- Creating a Memcached Cache Cluster (AWS CLI) (p. 84)

Creating a Memcached Cache Cluster (AWS CLI)

The following CLI code creates a Memcached cache cluster with 3 nodes.

For Linux, macOS, or Unix:

```bash
aws elasticache create-cache-cluster \
  --cache-cluster-id my-cluster \
  --cache-node-type cache.r4.large \
  --engine memcached \
  --engine-version 1.4.24 \
  --cache-parameter-group default.memcached1.4 \
  --num-cache-nodes 3
```

For Windows:

```bash
aws elasticache create-cache-cluster ^
  --cache-cluster-id my-cluster ^
  --cache-node-type cache.r4.large ^
  --engine memcached ^
  --engine-version 1.4.24 ^
  --cache-parameter-group default.memcached1.4 ^
  --num-cache-nodes 3
```
Creating a Cluster (ElastiCache API)

To create a cluster using the ElastiCache API, use the CreateCacheCluster action.

Important
As soon as your cluster becomes available, you're billed for each hour or partial hour that the cluster is active, even if you're not using it. To stop incurring charges for this cluster, you must delete it. See Deleting a Cluster (p. 105).

Topics
• Creating a Memcached Cache Cluster (ElastiCache API) (p. 85)

Creating a Memcached Cache Cluster (ElastiCache API)

The following code creates a Memcached cluster with 3 nodes (ElastiCache API).

Line breaks are added for ease of reading.

https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateCacheCluster
  &CacheClusterId=my-cluster
  &CacheNodeType=cache.r4.large
  &Engine=memcached
  &NumCacheNodes=3
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150508T220302Z
  &Version=2015-02-02
  &X-Amz-Algorithm=AWS4-HMAC-SHA256
  &X-Amz-Credential=<credential>
  &X-Amz-Date=20150508T220302Z
  &X-Amz-Expires=20150508T220302Z
  &X-Amz-SignedHeaders=Host
  &X-Amz-Signature=<signature>
Viewing a Cluster's Details

You can view detail information about one or more clusters using the ElastiCache console, AWS CLI, or ElastiCache API.

Topics

- Viewing a Cluster's Details (Console) (p. 86)
- Viewing a Cluster's Details (AWS CLI) (p. 88)
- Viewing a Cluster's Details (ElastiCache API) (p. 89)

Viewing a Cluster's Details (Console)

You can view the details of a Memcached cluster using the ElastiCache console, the AWS CLI for ElastiCache, or the ElastiCache API.

The following procedure details how to view the details of a Memcached cluster using the ElastiCache console.

To view a Memcached cluster's details

1. Sign in to the AWS Management Console and open the Amazon ElastiCache console at https://console.aws.amazon.com/elasticache/.
2. From the list in the upper-right corner, choose the AWS Region you are interested in.
3. In the ElastiCache console dashboard, choose Memcached. This will display a list of all your clusters that are running any version of Memcached.
4. To see details of a cluster, choose the box to the left of the cluster's name.
5. To view node information:
   a. Choose the cluster's name.
   b. Choose the Nodes tab.
   c. To view metrics on one or more nodes, choose the box to the left of the Node ID, and then choose the time range for the metrics from the Time range list. Selecting multiple nodes will generate overlay graphs.
Metrics over the last hour for two Memcached nodes
Viewing a Cluster’s Details (AWS CLI)

You can view the details for a cluster using the AWS CLI `describe-cache-clusters` command. If the `--cache-cluster-id` parameter is omitted, details for multiple clusters, up to `--max-items`, are returned. If the `--cache-cluster-id` parameter is included, details for the specified cluster are returned. You can limit the number of records returned with the `--max-items` parameter.

The following code lists the details for `my-cluster`.

```bash
code
aws elasticache describe-cache-clusters --cache-cluster-id my-cluster
```

The following code list the details for up to 25 clusters.

```bash
code
aws elasticache describe-cache-clusters --max-items 25
```

Example

For Linux, macOS, or Unix:

```bash
code
aws elasticache describe-cache-clusters \ 
  --cache-cluster-id my-cluster \ 
  --show-cache-node-info
```

For Windows:

```bash
code
aws elasticache describe-cache-clusters ^ 
  --cache-cluster-id my-cluster ^ 
  --show-cache-node-info
```

This operation produces output similar to the following (JSON format):

```json
code
{
  "CacheClusters": [
    {
      "Engine": "memcached",
      "CacheNodes": [
        {
          "CacheNodeId": "0001",
          "Endpoint": {
            "Port": 11211,
            "Address": "my-cluster.7ef-example.0001.usw2.cache.amazonaws.com"
          },
          "CacheNodeStatus": "available",
          "ParameterGroupStatus": "in-sync",
          "CustomerAvailabilityZone": "us-west-2b"
        },
        {
          "CacheNodeId": "0002",
          "Endpoint": {
            "Port": 11211,
            "Address": "my-cluster.7ef-example.0002.usw2.cache.amazonaws.com"
          },
          "CacheNodeStatus": "available",
          "ParameterGroupStatus": "in-sync",
          "CustomerAvailabilityZone": "us-west-2b"
        }
      ]
    }
  ]
}
```
Viewing a Cluster’s Details (ElastiCache API)

You can view the details for a cluster using the ElastiCache API DescribeCacheClusters action. If the CacheClusterId parameter is included, details for the specified cluster are returned. If the CacheClusterId parameter is omitted, details for up to MaxRecords (default 100) clusters are returned. The value for MaxRecords cannot be less than 20 or greater than 100.

The following code lists the details for my-cluster.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheClusters
&CacheClusterId=my-cluster
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
```
The following code lists the details for up to 25 clusters.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
  ?Action=DescribeCacheClusters
  &MaxRecords=25
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &X-Amz-Credential=<credential>
```

For more information, see the ElastiCache API reference topic [DescribeCacheClusters](https://docs.aws.amazon.com/elasticache/latest/memcachedguide/viewing-a-clusters-details.html).
Modifying an ElastiCache Cluster

In addition to adding or removing nodes from a cluster, there can be times where you need to make other changes to an existing cluster, such as, adding a security group, changing the maintenance window or a parameter group.

We recommend that you have your maintenance window fall at the time of lowest usage. Thus it might need modification from time to time.

When you change a cluster's parameters, the change is applied to the cluster either immediately or after the cluster is restarted. This is true whether you change the cluster's parameter group itself or a parameter value within the cluster's parameter group. To determine when a particular parameter change is applied, see the Changes Take Effect column in the tables for Memcached Specific Parameters (p. 151) and . For information on rebooting a cluster, see Rebooting a Cluster (p. 93).

Using the AWS Management Console

To modify a cluster (console)

2. From the list in the upper-right corner, choose the AWS Region where the cluster you want to modify is located.
3. In the navigation pane, choose the engine running on the cluster you want to modify.

A list of the chosen engine's clusters appears.
4. In the list of clusters, choose the name of the cluster, not the box to the left of the cluster's name, you want to modify.
5. Choose Modify.

The Modify Cluster window appears.
6. In the Modify Cluster window, make the modification(s) you want.

Important
You can upgrade to newer engine versions. For more information on doing so, see Upgrading Engine Versions (p. 36). However, you can't downgrade to older engine versions except by deleting the existing cluster and creating it again.

The Apply Immediately box applies only to engine version modifications. To apply changes immediately, choose the Apply Immediately check box. If this box is not chosen, engine version modifications will be applied during the next maintenance window. Other modifications, such as changing the maintenance window, are applied immediately.
7. Choose Modify.

Using the AWS CLI

You can modify an existing cluster using the AWS CLI modify-cache-cluster operation. To modify a cluster's configuration value, specify the cluster's ID, the parameter to change and the parameter's new value. The following example changes the maintenance window for a cluster named my-cluster and applies the change immediately.

Important
You can upgrade to newer engine versions. For more information on doing so, see Upgrading Engine Versions (p. 36). However, you can't downgrade to older engine versions except by deleting the existing cluster and creating it again.
For Linux, macOS, or Unix:

```
aws elasticache modify-cache-cluster \
  --cache-cluster-id my-cluster \
  --preferred-maintenance-window sun:23:00-mon:02:00
```

For Windows:

```
aws elasticache modify-cache-cluster ^
  --cache-cluster-id my-cluster ^
  --preferred-maintenance-window sun:23:00-mon:02:00
```

The `--apply-immediately` parameter applies only to modifications in node type, engine version, and changing the number of nodes in a cluster. If you want to apply any of these changes immediately, use the `--apply-immediately` parameter. If you prefer postponing these changes to your next maintenance window, use the `--no-apply-immediately` parameter. Other modifications, such as changing the maintenance window, are applied immediately.

For more information, see the AWS CLI for ElastiCache topic `modify-cache-cluster`.

**Using the ElastiCache API**

You can modify an existing cluster using the ElastiCache API `ModifyCacheCluster` operation. To modify a cluster's configuration value, specify the cluster's ID, the parameter to change and the parameter's new value. The following example changes the maintenance window for a cluster named `my-cluster` and applies the change immediately.

**Important**

You can upgrade to newer engine versions. For more information on doing so, see Upgrading Engine Versions (p. 36). However, you can't downgrade to older engine versions except by deleting the existing cluster and creating it again.

Line breaks are added for ease of reading.

```
https://elastcache.us-west-2.amazonaws.com/
  ?Action=ModifyCacheCluster
  &CacheClusterId=my-cluster
  &PreferredMaintenanceWindow=sun:23:00-mon:02:00
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150901T220302Z
  &X-Amz-Algorithm=AWS4-HMAC-SHA256
  &X-Amz-Date=20150202T220302Z
  &X-Amz-SignedHeaders=Host
  &X-Amz-Expires=20150901T220302Z
  &X-Amz-Credential=<credential>
  &X-Amz-Signature=<signature>
```

The `ApplyImmediately` parameter applies only to modifications in node type, engine version, and changing the number of nodes in a cluster. If you want to apply any of these changes immediately, set the `ApplyImmediately` parameter to `true`. If you prefer postponing these changes to your next maintenance window, set the `ApplyImmediately` parameter to `false`. Other modifications, such as changing the maintenance window, are applied immediately.

For more information, see the ElastiCache API reference topic `ModifyCacheCluster`.

---

API Version 2015-02-02

92
Rebooting a Cluster

Some changes require that the cluster be rebooted for the changes to be applied. For example, for some parameters, changing the parameter value in a parameter group is only applied after a reboot.

When you reboot a cluster, the cluster flushes all its data and restarts its engine. During this process you cannot access the cluster. Because the cluster flushed all its data, when the cluster is available again, you are starting with an empty cluster.

You are able to reboot a cluster using the ElastiCache console, the AWS CLI, or the ElastiCache API. Whether you use the ElastiCache console, the AWS CLI or the ElastiCache API, you can only initiate rebooting a single cluster. To reboot multiple clusters you must iterate on the process or operation.

Using the AWS Management Console

You can reboot a cluster using the ElastiCache console.

To reboot a cluster (console)

2. From the list in the upper-right corner, choose the AWS Region you are interested in.
3. In the navigation pane, choose the engine running on the cluster you want to reboot.
   A list of clusters running the chosen engine will appear.
4. Choose the cluster to reboot by choosing on the box to the left of the cluster's name.
   The Reboot button will become active.
   If you choose more than one cluster, the Reboot button becomes disabled.
5. Choose Reboot.
   The reboot cluster confirmation screen appears.
6. To reboot the cluster, choose Reboot. The status of the cluster will change to rebooting cluster nodes.
   To not reboot the cluster, choose Cancel.

To reboot multiple clusters, repeat steps 2 through 5 for each cluster you want to reboot. You do not need to wait for one cluster to finish rebooting to reboot another.

Using the AWS CLI

To reboot a cluster (AWS CLI), use the reboot-cache-cluster CLI operation.

To reboot specific nodes in the cluster, use the --cache-node-ids-to-reboot to list the specific clusters to reboot. The following command reboots the nodes 0001, 0002, and 0004 of my-cluster.

For Linux, macOS, or Unix:

```bash
aws elasticache reboot-cache-cluster
   --cache-cluster-id my-cluster
   --cache-node-ids-to-reboot 0001 0002 0004
```

For Windows:
To reboot all the nodes in the cluster, use the --cache-node-ids-to-reboot parameter and list all the cluster's node ids. For more information, see `reboot-cache-cluster`.

**Using the ElastiCache API**

To reboot a cluster using the ElastiCache API, use the `RebootCacheCluster` action.

To reboot specific nodes in the cluster, use the `CacheNodeIdsToReboot` to list the specific clusters to reboot. The following command reboots the nodes 0001, 0002, and 0004 of `my-cluster`.

https://elasticache.us-west-2.amazonaws.com/
?Action=RebootCacheCluster
&CacheClusterId=my-cluster
&CacheNodeIdsToReboot.member.1=0001
&CacheNodeIdsToReboot.member.2=0002
&CacheNodeIdsToReboot.member.3=0004
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>

To reboot all the nodes in the cluster, use the `CacheNodeIdsToReboot` parameter and list all the cluster's node ids. For more information, see `RebootCacheCluster`.

**Adding Nodes to a Cluster**

Adding nodes to a Memcached cluster increases the number of you cluster's partitions. When you change the number of partitions in a cluster some of your key spaces need to be remapped so that they are mapped to the right node. Remapping key spaces temporarily increases the number of cache misses on the cluster. For more information, see Configuring Your ElastiCache Client for Efficient Load Balancing (p. 31).

You can use the ElastiCache Management Console, the AWS CLI or ElastiCache API to add nodes to your cluster.

**Using the AWS Management Console**

**Topics**

- To add nodes to a cluster (console) (p. 94)

**To add nodes to a cluster (console)**

The following procedure can be used to add nodes to a cluster.

2. In the navigation pane, choose the engine running on the cluster you want to add nodes to.
   A list of clusters running the chosen engine appears.
3. From the list of clusters, choose the name of the cluster, not the box to the left of the cluster's name, you want to add a node to.

4. Choose Add node.

5. Complete the information requested in the Add Node dialog box.

6. Choose the Apply Immediately - Yes button to add this node immediately, or choose No to add this node during the cluster’s next maintenance window.

### Impact of New Add and Remove Requests on Pending Requests

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Pending Operation</th>
<th>New Request</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>Delete</td>
<td>Delete</td>
<td>The new delete request, pending or immediate, replaces the pending delete request. For example, if nodes 0001, 0003, and 0007 are pending deletion and a new request to delete nodes 0002 and 0004 is issued, only nodes 0002 and 0004 will be deleted. Nodes 0001, 0003, and 0007 will not be deleted.</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>Delete</td>
<td>Create</td>
<td>The new create request, pending or immediate, replaces the pending delete request. For example, if nodes 0001, 0003, and 0007 are pending deletion and a new request to create a node is issued, a new node will be created and nodes 0001, 0003, and 0007 will not be deleted.</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>Create</td>
<td>Delete</td>
<td>The new delete request, pending or immediate, replaces the pending create request. For example, if there is a pending request to create two nodes and a new request is issued to delete node 0003, no new nodes will be created and node 0003 will be deleted.</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>Create</td>
<td>Create</td>
<td>The new create request is added to the pending create request. For example, if there is a pending request to create two nodes and a new request is issued to create three nodes, the new requests is added to the pending request and five nodes will be created.</td>
</tr>
</tbody>
</table>

**Important**

If the new create request is set to Apply Immediately - Yes, all create requests are performed immediately. If the new create request is set to Apply Immediately - No, all create requests are pending.

To determine what operations are pending, choose the Description tab and check to see how many pending creations or deletions are shown. You cannot have both pending creations and pending deletions.
7. Choose the **Add** button.

After a few moments, the new nodes should appear in the nodes list with a status of **creating**. If they don't appear, refresh your browser page. When the node's status changes to **available** the new node is able to be used.

### Using the AWS CLI

To add nodes to a cluster using the AWS CLI, use the AWS CLI operation `modify-cache-cluster` with the following parameters:

- `--cache-cluster-id` The ID of the cache cluster you want to add nodes to.
- `--num-cache-nodes` The `--num-cache-nodes` parameter specifies the number of nodes you want in this cluster after the modification is applied. To add nodes to this cluster, `--num-cache-nodes` must be greater than the current number of nodes in this cluster. If this value is less than the current number of nodes, ElastiCache expects the parameter `cache-node-ids-to-remove` and a list of nodes to remove from the cluster. For more information, see Using the AWS CLI (p. 101).
- `--apply-immediately` or `--no-apply-immediately` which specifies whether to add these nodes immediately or at the next maintenance window.

For Linux, macOS, or Unix:

```
aws elasticache modify-cache-cluster \
  --cache-cluster-id my-cluster \
  --num-cache-nodes 5 \
  --apply-immediately
```

For Windows:

```
aws elasticache modify-cache-cluster ^
  --cache-cluster-id my-cluster ^
  --num-cache-nodes 5 ^
  --apply-immediately
```
Adding Nodes to a Cluster

This operation produces output similar to the following (JSON format):

```
{
  "CacheCluster": {
    "Engine": "memcached",
    "CacheParameterGroup": {
      "CacheNodeIdsToReboot": [],
      "CacheParameterGroupName": "default.memcached1.4",
      "ParameterApplyStatus": "in-sync"
    },
    "CacheClusterId": "my-cluster",
    "PreferredAvailabilityZone": "us-west-2b",
    "ConfigurationEndpoint": {
      "Port": 11211,
      "Address": "rlh-mem000.7alc7bf-example.cfg.usw2.cache.amazonaws.com"
    },
    "CacheSecurityGroups": [],
    "AutoMinorVersionUpgrade": true,
    "CacheClusterStatus": "modifying",
    "NumCacheNodes": 2,
    "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
    "SecurityGroups": [
      {
        "Status": "active",
        "SecurityGroupId": "sg-dbe93fa2"
      }
    ],
    "CacheSubnetGroupName": "default",
    "EngineVersion": "1.4.24",
    "PendingModifiedValues": {
      "NumCacheNodes": 5
    },
    "PreferredMaintenanceWindow": "sat:09:00-sat:10:00",
    "CacheNodeType": "cache.m3.medium"
  }
}
```

For more information, see the AWS CLI topic modify-cache-cluster.

Using the ElastiCache API

To add nodes to a cluster (ElastiCache API)

- Call the ModifyCacheCluster API operation with the following parameters:
  - **CacheClusterId** The ID of the cluster you want to add nodes to.
  - **NumCacheNodes** The NumCacheNodes parameter specifies the number of nodes you want in this cluster after the modification is applied. To add nodes to this cluster, NumCacheNodes must be greater than the current number of nodes in this cluster. If this value is less than the current number of nodes, ElastiCache expects the parameter CacheNodeIdsToRemove with a list of nodes to remove from the cluster (see Using the ElastiCache API (p. 102)).
  - **ApplyImmediately** Specifies whether to add these nodes immediately or at the next maintenance window.
  - **Region** Specifies the AWS region of the cluster you want to add nodes to.

The following example shows a call to add nodes to a cluster.
Example

```
https://elasticache.us-west-2.amazonaws.com/
?Action=ModifyCacheCluster
&ApplyImmediately=true
&NumCacheNodes=5
&CacheClusterId=my-cluster
&Region=us-east-2
&Version=2014-12-01
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&X-Amz-Algorithm=AWS4-HMAC-SHA256
&X-Amz-Date=20141201T220302Z
&X-Amz-SignedHeaders=Host
&X-Amz-Expires=20141201T220302Z
&X-Amz-Credential=<credential>
&X-Amz-Signature=<signature>
```

For more information, see ElastiCache API topic [ModifyCacheCluster](https://elasticache.us-west-2.amazonaws.com/).
Removing Nodes from a Cluster

Each time you change the number of nodes in a Memcached cluster, you must re-map at least some of your keyspace so it maps to the correct node. For more detailed information on load balancing a Memcached cluster, see Configuring Your ElastiCache Client for Efficient Load Balancing (p. 31).

You can delete a node from a cluster using the AWS Management Console, the AWS CLI, or the ElastiCache API.

Topics

- Using the AWS Management Console (p. 99)
- Using the AWS CLI (p. 101)
- Using the ElastiCache API (p. 102)

Using the AWS Management Console

To remove nodes from a cluster (console)

2. From the list in the upper-right corner, choose the AWS Region of the cluster you want to remove nodes from.
3. In the navigation pane, choose the engine running on the cluster you want to remove a node.
   A list of clusters running the chosen engine appears.
4. From the list of clusters, choose the cluster name from which you want to remove a node.
   A list of the cluster’s nodes appears.
5. Choose the box to the left of the node ID for the node you want to remove. Using the ElastiCache console, you can only delete one node at a time, so choosing multiple nodes will disable the Delete node button.
   The Delete Node dialog appears.
6. To delete the node, complete the Delete Node dialog box and choose Delete Node. To not delete the node, choose the Cancel.

Impact of New Add and Remove Requests on Pending Requests

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Pending Operation</th>
<th>New Request</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>Delete</td>
<td>Delete</td>
<td>The new delete request, pending or immediate, replaces the pending delete request. For example, if nodes 0001, 0003, and 0007 are pending deletion and a new request to delete nodes 0002 and 0004 is issued, only nodes 0002 and 0004 will be deleted. Nodes 0001, 0003, and 0007 will not be deleted.</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>Delete</td>
<td>Create</td>
<td>The new create request, pending or immediate, replaces the pending delete request. For example, if nodes 0001, 0003, and 0007 are pending deletion and a new request to create a node is issued,</td>
</tr>
</tbody>
</table>
### Removing Nodes from a Cluster

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Pending Operation</th>
<th>New Request</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>a new node will be created and nodes 0001, 0003, and 0007 will not be deleted.</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>Create</td>
<td>Delete</td>
<td>The new delete request, pending or immediate, replaces the pending create request. For example, if there is a pending request to create two nodes and a new request is issued to delete node 0003, no new nodes will be created and node 0003 will be deleted.</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>Create</td>
<td>Create</td>
<td>The new create request is added to the pending create request. For example, if there is a pending request to create two nodes and a new request is issued to create three nodes, the new requests is added to the pending request and five nodes will be created. <strong>Important</strong> If the new create request is set to <strong>Apply Immediately - Yes</strong>, all create requests are performed immediately. If the new create request is set to <strong>Apply Immediately - No</strong>, all create requests are pending.</td>
</tr>
</tbody>
</table>

To determine what operations are pending, choose the **Description** tab and check to see how many pending creations or deletions are shown. You cannot have both pending creations and pending deletions.
Using the AWS CLI

1. Identify the IDs of the nodes you want to remove. For more information, see Viewing a Cluster's Details (p. 86).
2. Use the modify-cache-cluster CLI operation with a list of the nodes to remove, as in the following example.

   To remove nodes from a cluster using the command-line interface, use the command modify-cache-cluster with the following parameters:

   • --cache-cluster-id The ID of the cache cluster you want to remove nodes from.
   • --num-cache-nodes The --num-cache-nodes parameter specifies the number of nodes you want in this cluster after the modification is applied.
   • --cache-node-ids-to-remove A list of node IDs you want removed from this cluster.
   • --apply-immediately or --no-apply-immediately Specifies whether to remove these nodes immediately or at the next maintenance window.
   • --region Specifies the region of the cluster you want to remove nodes from.

   The following example immediately removes node 0001 from the cluster my-cluster.

   For Linux, macOS, or Unix:

   ```
   aws elasticache modify-cache-cluster \
   --cache-cluster-id my-cluster \
   --num-cache-nodes 2 \
   --cache-node-ids-to-remove 0001 \
   --region us-east-2 \
   --apply-immediately
   ```

   For Windows:

   ```
   aws elasticache modify-cache-cluster ^
   --cache-cluster-id my-cluster ^
   --num-cache-nodes 2 ^
   --cache-node-ids-to-remove 0001 ^
   --region us-east-2 ^
   --apply-immediately
   ```

   This operation produces output similar to the following (JSON format):

   ```
   {
   "CacheCluster": {
   "Engine": "memcached",
   "CacheParameterGroup": {
   "CacheNodeIdsToReboot": [],
   "CacheParameterGroupName": "default.memcached1.4",
   "ParameterApplyStatus": "in-sync"
   },
   "CacheClusterId": "my-cluster",
   "PreferredAvailabilityZone": "us-east-2b",
   "ConfigurationEndpoint": {
   "Port": 11211,
   "Address": "rlh-mem000.7ef-example.cfg.usw2.cache.amazonaws.com"
   },
   "CacheSecurityGroups": [],
   "AutoMinorVersionUpgrade": true,
   "API Version": "2015-02-02"
   }
   ```
Removing Nodes from a Cluster

For more information, see the AWS CLI topics `describe-cache-cluster` and `modify-cache-cluster`.

Using the ElastiCache API

To remove nodes using the ElastiCache API, call the `ModifyCacheCluster` API operation with the cache cluster ID and a list of nodes to remove, as shown:

- `CacheClusterId` The ID of the cache cluster you want to remove nodes from.
- `NumCacheNodes` The `NumCacheNodes` parameter specifies the number of nodes you want in this cluster after the modification is applied.
- `CacheNodeIdsToRemove.member.n` The list of node IDs to remove from the cluster.
  - `CacheNodeIdsToRemove.member.1=0004`
  - `CacheNodeIdsToRemove.member.1=0005`
- `ApplyImmediately` Specifies whether to remove these nodes immediately or at the next maintenance window.
- `Region` Specifies the region of the cluster you want to remove a node from.

The following example immediately removes nodes 0004 and 0005 from the cluster my-cluster.

https://elasticache.us-west-2.amazonaws.com/
?Action=ModifyCacheCluster
&CacheClusterId=my-cluster
&ApplyImmediately=true
&CacheNodeIdsToRemove.member.1=0004
&CacheNodeIdsToRemove.member.2=0005
&NumCacheNodes=3
&Region=us-east-2
&Version=2014-12-01
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&X-Amz-Algorithm=AWS4-HMAC-SHA256
&X-Amz-Date=20141201T220302Z
&X-Amz-SignedHeaders=Host
Removing Nodes from a Cluster

For more information, see ElastiCache API topic ModifyCacheCluster.
Canceling Pending Add or Delete Node Operations

If you elected to not apply a change immediately, the operation has pending status until it is performed at your next maintenance window. You can cancel any pending operation.

To cancel a pending operation

1. Sign in to the AWS Management Console and open the ElastiCache console at https://console.aws.amazon.com/elasticache/
2. From the list in the upper-right corner, choose the AWS Region you want to cancel a pending add or delete node operation in.
3. In the navigation pane, choose the engine running on the cluster that has pending operations you want to cancel. A list of clusters running the chosen engine will appear.
4. In the list of clusters, choose the name of the cluster, not the box to the left of the cluster’s name, that has pending operations you want to cancel.
5. To determine what operations are pending, choose the Description tab and check to see how many pending creations or deletions are shown. You cannot have both pending creations and pending deletions.
6. Choose the Nodes tab.
7. To cancel all pending operations, click Cancel Pending. The Cancel Pending dialog box appears.
8. Confirm that you want to cancel all pending operations by choosing the Cancel Pending button, or to keep the operations, choose Cancel.
Deleting a Cluster

As long as a cluster is in the available state, you are being charged for it, whether or not you are actively using it. To stop incurring charges, delete the cluster.

**Warning**

When you delete an ElastiCache for Redis cluster, your manual snapshots are retained. You will also have an option to create a final snapshot before the cluster is deleted. Automatic cache snapshots are not retained.

Using the AWS Management Console

The following procedure deletes a single cluster from your deployment. To delete multiple clusters, repeat the procedure for each cluster you want to delete. You do not need to wait for one cluster to finish deleting before starting the procedure to delete another cluster.

**To delete a cluster**

1. Sign in to the AWS Management Console and open the Amazon ElastiCache console at https://console.aws.amazon.com/elasticache/.
2. In the ElastiCache console dashboard, select the engine the cluster you want to delete is running. A list of all clusters running the selected engine appears.
3. To select the cluster to delete, select the cluster’s name from the list of clusters.

   **Important**
   
   You can only delete one cluster at a time from the ElastiCache console. Selecting multiple clusters disables the delete operation.

4. Select the Actions button and then select Delete from the list of actions.
5. In the Delete Cluster confirmation screen, choose Delete to delete the cluster, or select Cancel to keep the cluster.

   If you chose Delete, the status of the cluster changes to deleting.

As soon as your cluster is no longer listed in the list of clusters, you stop incurring charges for it.

Using the AWS CLI

The following code deletes the cache cluster my-cluster.

```
aws elasticache delete-cache-cluster --cache-cluster-id my-cluster
```

The delete-cache-cluster CLI action only deletes one cache cluster. To delete multiple cache clusters, call delete-cache-cluster for each cache cluster you want to delete. You do not need to wait for one cache cluster to finish deleting before deleting another.

For Linux, macOS, or Unix:

```
aws elasticache delete-cache-cluster --cache-cluster-id my-cluster --region us-east-2
```

For Windows:

```
aws elasticache delete-cache-cluster
```
Deleting a Cluster

```bash
--cache-cluster-id my-cluster
--region us-east-2
```

For more information, see the AWS CLI for ElastiCache topic `delete-cache-cluster`.

**Using the ElastiCache API**

The following code deletes the cluster `my-cluster`.

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=DeleteCacheCluster
  &CacheClusterId=my-cluster
  &Region us-east-2
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T220302Z
  &X-Amz-Algorithm=AWS4-HMAC-SHA256
  &X-Amz-Date=20150202T220302Z
  &X-Amz-SignedHeaders=Host
  &X-Amz-Expires=20150202T220302Z
  &X-Amz-Credential=<credential>
  &X-Amz-Signature=<signature>
```

The `DeleteCacheCluster` API operation only deletes one cache cluster. To delete multiple cache clusters, call `DeleteCacheCluster` for each cache cluster you want to delete. You do not need to wait for one cache cluster to finish deleting before deleting another.

For more information, see the ElastiCache API reference topic `DeleteCacheCluster`.

API Version 2015-02-02
Accessing Your Cluster

Your Amazon ElastiCache instances are designed to be accessed through an Amazon EC2 instance.

If you launched your ElastiCache instance in an Amazon Virtual Private Cloud (Amazon VPC), you can access your ElastiCache instance from an Amazon EC2 instance in the same Amazon VPC. Or, by using VPC peering, you can access your ElastiCache instance from an Amazon EC2 in a different Amazon VPC.

If you launched your ElastiCache instance in EC2 Classic, you allow the EC2 instance to access your cluster by granting the Amazon EC2 security group associated with the instance access to your cache security group. By default, access to a cluster is restricted to the account that launched the cluster.

Topics
- Determine the Cluster's Platform (p. 107)
- Grant Access to Your Cluster (p. 109)

Determine the Cluster's Platform

Before you continue, determine whether you launched your cluster into EC2-VPC or EC2-Classic.

For more information, see Detecting Your Supported Platforms and Whether You Have a Default VPC.

Determining Your Clusters Platform using the ElastiCache Console

The following procedure uses the ElastiCache console to determine whether you launched your cluster into EC2-VPC or EC2-Classic.

To determine a cluster's platform using the ElastiCache console

2. To see a list of your clusters running the Memcached engine, in the left navigation pane, choose Memcached.
3. In the list of clusters, expand the cluster you want to authorize access to by choosing the box to the left of the cluster name.
4. Locate Subnet group:

   - If the Subnet group has a name, as shown here, you launched your cluster in EC2-VPC and should continue at You Launched Your Cluster into EC2-VPC (p. 109).
   - If there is a dash (-) instead of a Subnet group name, you launched your cluster in EC2-Classic and should continue at You Launched Your Cluster Running in EC2-Classic (p. 109).

For more information, see Detecting Your Supported Platforms and Whether You Have a Default VPC.
Determining Your Clusters Platform using the AWS CLI

The following procedure uses the AWS CLI to determine whether you launched your cluster into EC2-VPC or EC2-Classic.

To determine a cluster's platform using the AWS CLI

1. Open a command window.
2. At the command prompt, run the following command.

For Linux, macOS, or Unix:

```bash
aws elasticache describe-cache-clusters \
  --show-cache-cluster-details \
  --cache-cluster-id my-cluster
```

For Windows:

```bash
aws elasticache describe-cache-clusters ^
  --show-cache-cluster-details ^
  --cache-cluster-id my-cluster
```

JSON output from this command will look something like this. Some of the output is omitted to save space.

```json
{
  "CacheClusters": [
    {
      "Engine": "memcached",
      "AuthTokenEnabled": false,
      "CacheParameterGroup": {
        "CacheNodeIdsToReboot": [],
        "CacheParameterGroupName": "default.memcached1.4",
        "ParameterApplyStatus": "in-sync"
      },
      "CacheClusterId": "my-cluster-001",
      "CacheSecurityGroups": [],
      "NumCacheNodes": 1,
      "AtRestEncryptionEnabled": false,
      "CacheClusterCreateTime": "2018-01-16T20:09:34.449Z",
      "ReplicationGroupId": "my-cluster",
      "AutoMinorVersionUpgrade": true,
      "CacheClusterStatus": "available",
      "PreferredAvailabilityZone": "us-east-2a",
      "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
      "SecurityGroups": [
        {
          "Status": "active",
          "SecurityGroupId": "sg-e8c03081"
        }
      ],
      "TransitEncryptionEnabled": false,
      "CacheSubnetGroupName": "default",
      "EngineVersion": "3.2.10",
      "PendingModifiedValues": {},
      "PreferredMaintenanceWindow": "sat:05:30-sat:06:30",
      "CacheNodeType": "cache.t2.medium"
    }
  ]
}
```
• If there is a value for `CacheSubnetGroupName`, you launched your cluster in EC2-VPC and should continue at You Launched Your Cluster into EC2-VPC (p. 109).

• If there is no value for `CacheSubnetGroupName`, you launched your cluster in EC2-Classic and should continue at You Launched Your Cluster Running in EC2-Classic (p. 109).

Grant Access to Your Cluster

You Launched Your Cluster into EC2-VPC

If you launched your cluster into an Amazon Virtual Private Cloud (Amazon VPC), you can connect to your ElastiCache cluster only from an Amazon EC2 instance that is running in the same Amazon VPC. In this case, you will need to grant network ingress to the cluster.

To grant network ingress from an Amazon VPC security group to a cluster

1. Sign in to the AWS Management Console and open the Amazon EC2 console at https://console.aws.amazon.com/ec2/.
3. From the list of security groups, choose the security group for your Amazon VPC. Unless you created a security group for ElastiCache use, this security group will be named default.
4. Choose the Inbound tab, and then do the following:
   a. Choose Edit.
   b. Choose Add rule.
   c. In the Type column, choose Custom TCP rule.
   d. In the Port range box, type the port number for your cluster node. This number must be the same one that you specified when you launched the cluster. The default port for Memcached is 11211.
   e. In the Source box, choose Anywhere which has the port range (0.0.0.0/0) so that any Amazon EC2 instance that you launch within your Amazon VPC can connect to your ElastiCache nodes.

   Important
   Opening up the ElastiCache cluster to 0.0.0.0/0 (Step 4.e.) does not expose the cluster to the Internet because it has no public IP address and therefore cannot be accessed from outside the VPC. However, the default security group may be applied to other Amazon EC2 instances in the customer's account, and those instances may have a public IP address. If they happen to be running something on port 6379, then that service could be exposed unintentionally. Therefore, we recommend creating a VPC Security Group that will be used exclusively by ElastiCache. For more information, see Custom Security Groups.
   f. Choose Save.

When you launch an Amazon EC2 instance into your Amazon VPC, that instance will be able to connect to your ElastiCache cluster.

You Launched Your Cluster Running in EC2-Classic

If you launched your cluster into EC2-Classic, to allow an Amazon EC2 instance to access your cluster you will need to grant the Amazon EC2 security group associated with the instance access to your cache security group.
To grant an Amazon EC2 security group access to a cluster

2. To see a list of security groups, from the left navigation pane, choose Security Groups.  
   Important
   If Security Groups is not listed in the navigation pane, you launched your cluster in EC2-VPC rather than EC2-Classic and should follow the instructions at You Launched Your Cluster into EC2-VPC (p. 109).
3. Choose the box to the left of default security group.
4. From the list at the bottom of the screen, choose the EC2 Security Group Name you want to authorize.
5. To authorize access, choose Add.

   Amazon EC2 instances that are associated with the security group are now authorized to connect to your ElastiCache cluster.

To revoke a security group's access, locate the security group in the list of authorized security groups, and then choose Remove.

For more information on ElastiCache Security Groups, see Security Groups: EC2-Classic (p. 222).
Accessing ElastiCache Resources from Outside AWS

ElastiCache is a service designed to be used internally to your VPC. External access is discouraged due to the latency of Internet traffic and security concerns. However, if external access to ElastiCache is required for test or development purposes, it can be done through a VPN.

Using the AWS Client VPN, you allow external access to your ElastiCache nodes with the following benefits:

• Restricted access to approved users or authentication keys;
• Encrypted traffic between the VPN Client and the AWS VPN endpoint;
• Limited access to specific subnets or nodes;
• Easy revocation of access from users or authentication keys;
• Audit connections;

The following procedures demonstrate how to:

Topics
- Create a Certificate Authority (p. 111)
- Configuring AWS Client VPN Components (p. 112)
- Configure the VPN Client (p. 114)

Create a Certificate Authority

It is possible to create a Certificate Authority (CA) using different techniques or tools. We suggest the easy-rsa utility, provided by the OpenVPN project. Regardless of the option you choose, make sure to keep the keys secure. The following procedure downloads the easy-rsa scripts, creates the Certificate Authority and the keys to authenticate the first VPN client:

• To create the initial certificates, open a terminal and do the following:
  - git clone https://github.com/OpenVPN/easy-rs
  - cd easy-rsa
  - ./easyrsa3/easyrsa init-pki
  - ./easyrsa3/easyrsa build-ca nopass
  - ./easyrsa3/easyrsa build-server-full server nopass
  - ./easyrsa3/easyrsa build-client-full client1.domain.tld nopass

A pki subdirectory containing the certificates will be created under easy-rsa.

• Submit the server certificate to the AWS Certificate manager (ACM):
  - On the ACM console, select Certificate Manager.
  - Select Import Certificate.
  - Enter the public key certificate available in the easy-rsa/pki/issued/server.crt file in the Certificate body field.
  - Paste the private key available in the easy-rsa/pki/private/server.key in the Certificate private key field. Make sure to select all the lines between BEGIN AND END PRIVATE KEY (including the BEGIN and END lines).
  - Paste the CA public key available on the easy-rsa/pki/ca.crt file in the Certificate chain field.
  - Select Review and import.
  - Select Import.
To submit the server’s certificates to ACM using the AWS CLI, run the following command:

```bash
aws acm import-certificate --certificate file://easy-rsa/pki/issued/server.crt
--private-key file://easy-rsa/pki/private/server.key --certificate-chain
file://easy-rsa/pki/ca.crt --region region
```

Note the Certificate ARN for future use.

**Configuring AWS Client VPN Components**

**Using the AWS Console**

On the AWS console, select Services and then VPC.

Under Virtual Private Network, select Client VPN Endpoints and do the following:

**Configuring AWS Client VPN components**

- Select Create Client VPN Endpoint.
- Specify the following options:
  - **Client IPv4 CIDR**: use a private network with a netmask of at least /22 range. Make sure that the selected subnet does not conflict with the VPC networks’ addresses. Example: 10.0.0.0/22.
  - **In Server certificate ARN**, select the ARN of the certificate previously imported.
  - Select Use mutual authentication.
  - **In Client certificate ARN**, select the ARN of the certificate previously imported.
  - Select Create Client VPN Endpoint.

**Using the AWS CLI**

Run the following command:

```bash
```

Example output:

```
"ClientVpnEndpointId": "cvpn-endpoint-0123456789abcdefg",
"Status": { "Code": "pending-associate" }, "DnsName": "cvpn-endpoint-0123456789abcdefg.prod.clientvpn.us-east-1.amazonaws.com" }
```

**Associate the target networks to the VPN endpoint**

- Select the new VPN endpoint, and then select the Associations tab.
- Select Associate and specify the following options.
  - **VPC**: Select the Elasticache Cluster’s VPC.
  - Select one of the Elasticache cluster’s networks. If in doubt, review the networks in the **Subnet Groups** on the Elasticache dashboard.
  - Select Associate. If necessary, repeat the steps for the remaining networks.

**Using the AWS CLI**

Run the following command:
aws ec2 associate-client-vpn-target-network --client-vpn-endpoint-id cvpn-endpoint-0123456789abcdefg --subnet-id subnet-0123456789abdcdef

Example output:
"Status": { "Code": "associating" }, "AssociationId": "cvpn-assoc-0123456789abdcdef" }

**Review the VPN security group**

The VPN Endpoint will automatically adopt the VPC's default security group. Check the inbound and outbound rules and confirm if the security group allows the traffic from the VPN network (defined on the VPN Endpoint settings) to the ElastiCache networks on the service ports (by default, 6379 for Redis and 11211 for Memcached).

If you need to change the security group assigned to the VPN Endpoint, proceed as follows:

- Select the current security group.
- Select **Apply Security Group**.
- Select the new Security Group.

**Using the AWS CLI**

Run the following command:

aws ec2 apply-security-groups-to-client-vpn-target-network --client-vpn-endpoint-id cvpn-endpoint-0123456789abcdefg --vpc-id vpc-0123456789abdcdef --security-group-ids sg-0123456789abdcdef

Example output:
"SecurityGroupIds": [ "sg-0123456789abdcdef" ] }

**Note**

The ElastiCache security group also needs to allow traffic coming from the VPN clients. The clients' addresses will be masked with the VPN Endpoint address, according to the VPC Network. Therefore, consider the VPC network (not the VPN Clients' network) when creating the inbound rule on the ElastiCache security group.

**Authorize the VPN access to the destination networks**

On the **Authorization** tab, select **Authorize Ingress** and specify the following:

- Destination network to enable access: Either use 0.0.0.0/0 to allow access to any network (including the Internet) or restrict the the ElastiCache networks/hosts.
- Under **Grant access to:**; select **Allow access to all users**.
- Select **Add Authorization Rules**.

**Using the AWS CLI**

Run the following command:

aws ec2 authorize-client-vpn-ingress --client-vpn-endpoint-id cvpn-endpoint-0123456789abcdefg --target-network-cidr 0.0.0.0/0 --authorize-all-groups

Example output:
{ "Status": { "Code": "authorizing" } }
Allowing access to the Internet from the VPN clients

If you need to browse the Internet through the VPN, you need to create an additional route. Select the Route Table tab and then select Create Route:

- Route destination: 0.0.0.0/0
- Target VPC Subnet ID: Select one of the associated subnets with access to the Internet.
- Select Create Route.

Using the AWS CLI

Run the following command:

```
aws ec2 create-client-vpn-route --client-vpn-endpoint-id cvpn-endpoint-0123456789abcdefg --destination-cidr-block 0.0.0.0/0 --target-vpc-subnet-id subnet-0123456789abdcdef
```

Example output:

```
{ "Status": { "Code": "creating" } }
```

Configure the VPN Client

On the AWS Client VPN Dashboard, select the VPN endpoint recently created and select Download Client Configuration. Copy the configuration file, and the files easy-rsa/pki/issued/client1.domain.tld.crt and easy-rsa/pki/private/client1.domain.tld.key. Edit the configuration file and change or add the following parameters:

- cert: add a new line with the parameter cert pointing to the client1.domain.tld.crt file. Use the full path to the file. Example: `cert /home/user/.cert/client1.domain.tld.crt`
- cert: key: add a new line with the parameter key pointing to the client1.domain.tld.key file. Use the full path to the file. Example: `key /home/user/.cert/client1.domain.tld.key`

Establish the VPN connection with the command: `sudo openvpn --config downloaded-client-config.ovpn`

Revoking access

If you need to invalidate the access from a particular client key, the key needs to be revoked in the CA. Then submit the revocation list to AWS Client VPN.

Revoking the key with easy-rsa:

- cd easy-rsa
- ./easyrsa3/easyrsa revoke client1.domain.tld
- Enter "yes" to continue, or any other input to abort.

```
Continue with revocation: `yes` ... * `./easyrsa3/easyrsa gen-crl
```

An updated CRL has been created. CRL file: `/home/user/easy-rsa/pki/crl.pem`

Importing the revocation list to the AWS Client VPN:

- On the AWS Management Console, select Services and then VPC.
- Select Client VPN Endpoints.
- Select the Client VPN Endpoint and then select Actions -> Import Client Certificate CRL.
- Paste the contents of the crl.pem file.
Replication Across AWS Regions Using Global Datastore

By using the Global Datastore for Redis feature, you can work with fully managed, fast, reliable, and secure replication across AWS Regions. Using this feature, you can create cross-Region read replica clusters for ElastiCache for Redis to enable low-latency reads and disaster recovery across AWS Regions.

In the following sections, you can find a description of how to work with global datastores.

Topics

- Overview (p. 115)
- Prerequisites and Limitations (p. 116)
- Using Global Datastores (Console) (p. 116)
- Using Global Datastores (CLI) (p. 124)

Overview

Each global datastore is a collection of one or more clusters that replicate to one another.

A global datastore consists of the following:

- **Primary (active) cluster** – A primary cluster accepts writes that are replicated to all clusters within the global datastore. A primary cluster also accepts read requests.
- **Secondary (passive) cluster** – A secondary cluster only accepts read requests and replicates data updates from a primary cluster. A secondary cluster needs to be in a different AWS Region than the primary cluster.

When you create a global datastore in ElastiCache, ElastiCache for Redis automatically replicates your data from the primary cluster to the secondary cluster. You choose the AWS Region where the Redis data should be replicated and then create a secondary cluster in that AWS Region. ElastiCache then sets up and manages automatic, asynchronous replication of data between the two clusters.

Using a global datastore for Redis provides the following advantages:

- **Geolocal performance** – By setting up remote replica clusters in additional AWS Regions and synchronizing your data between them, you can reduce latency of data access in that AWS Region. A global datastore can help increase the responsiveness of your application by serving low-latency, geolocal reads across AWS Regions.
- **Disaster recovery** – If your primary cluster in a global datastore experiences degradation, you can promote a secondary cluster as your new primary cluster. You can do so by connecting to any AWS Region that contains a secondary cluster.

Run the following command:

```
aws ec2 import-client-vpn-client-certificate-revocation-list --certificate-revocation-list file://./easy-rsa/pki/crl.pem --client-vpn-endpoint-id cvpn-endpoint-0123456789abcdefg
```

Example output:

```
Example output: { "Return": true }
```
The following diagram shows how global datastores can work.

Prerequisites and Limitations

Before getting started with global datastores, be aware of the following:

- Global datastores are supported in the following AWS Regions: ap-northeast-2, ap-southeast-1, ap-southeast-2, ap-northeast-1, eu-central-1, eu-west-2, eu-west-1, us-east-1, us-east-2, us-west-1, us-west-2.
- To use global datastores, use Redis engine version 5.0.6 or higher and R5 or M5 node types.
- All clusters—primary and secondary—in your global datastore should have the same number of master nodes, node type, engine version, and number of shards (in case of cluster-mode enabled). Each cluster in your global datastore can have a different number of read replicas to accommodate the read traffic local to that cluster.
- Replication must be enabled if you plan to use an existing single-node cluster.
- You can set up replication for a primary cluster from one AWS Region to a secondary cluster in up to two other AWS Regions.
- You can work with global datastores only in VPC clusters. For more information, see Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 204). Global datastores aren't supported when you use EC2-Classic.
- ElastiCache doesn't support autofailover from one AWS Region to another. When needed, you can promote a secondary cluster manually. For an example, see Promoting the Secondary Cluster to Primary (p. 123).
- To bootstrap from existing data, use an existing cluster as primary to create a global datastore. We don't support adding an existing cluster as secondary. The process of adding the cluster as secondary wipes data, which may result in data loss.
- Parameter updates are applied to all clusters when you modify a local parameter group of a cluster belonging to a global datastore.
- You can scale regional clusters both vertically (scaling up and down) and horizontally (scaling in and out). You can scale the clusters by modifying the global datastore. All the regional clusters in the global datastore are then scaled without interruption.
- Global datastores support customer master keys, encryption at rest, encryption in transit, and Redis AUTH. Security for cross-Region communication is provided through VPC peering.

Note
Global datastores support pub/sub messaging with the following stipulations:

- For cluster-mode disabled, pub/sub is fully supported. Events published on the master of the primary AWS Region are propagated to secondary AWS Regions.
- For cluster mode enabled, the following applies:
  - For published events that aren't in a keyspace, only subscribers in the same AWS Region receive the events.
  - For published keyspace event, subscribers in all AWS Regions receive the events.

Using Global Datastores (Console)

To create a global datastore using the console, follow this two-step process:

1. Create a primary cluster, either by using an existing cluster or creating a new cluster. The engine must be Redis 5.0.6 or later.
2. Add up to two secondary clusters in different AWS Regions, again using the Redis 5.0.6 engine or later.

The following procedures guide you on how to create a global datastore for Redis and perform other operations using the ElastiCache for Redis console.

Topics
- Creating a Global Datastore Using an Existing Cluster (p. 117)
- Creating a New Global Datastore Using a New Primary Cluster (p. 118)
- Viewing Global Datastore Details (p. 120)
- Adding a Region to a Global Datastore (p. 121)
- Modifying a Global Datastore (p. 122)
- Promoting the Secondary Cluster to Primary (p. 123)
- Removing a Region from a Global Datastore (p. 123)
- Deleting a Global Datastore (p. 124)

Creating a Global Datastore Using an Existing Cluster

In this scenario, you use an existing cluster to serve as the primary of the new global datastore. You then create a secondary, read-only cluster in a separate AWS Region. This secondary cluster receives automatic and asynchronous updates from the primary cluster.

Important
The existing cluster must use the Redis 5.0.6 engine or later.

To create a global datastore using an existing cluster

2. On the navigation pane, choose Redis and then choose a cluster.
3. For Actions, choose Setup Global Datastore.
4. On the Setup Global Datastore page, do the following:
   - Enter a value for Global Datastore Name suffix: This suffix is used to generate a unique name for the global datastore. You can search for the global datastore by using the suffix that you specify here.
   - (Optional) Enter a Description value.
5. Under Secondary cluster details, choose a different AWS Region where the cluster will be stored.
6. Under Redis settings, enter a value for Name and, optionally, for Description for the cluster.
7. Keep the following options as they are. They're prepopulated to match the primary cluster configuration, you can't change them.
   - Engine version
   - Node type
   - Parameter group

Note
ElastiCache autogenerates a new parameter group from values of the provided parameter group and applies the new parameter group to the cluster. Use this new parameter group to modify parameters on a global datastore. Each autogenerated parameter group is associated with one and only one cluster and, therefore, only one global datastore.

- Number of shards
- Encryption at rest
Note
You can supply a different encryption key by choosing Customer Managed Customer Master Key and selecting the key.

- Encryption in transit
- Redis AUTH

8. (Optional) As needed, update the remaining secondary cluster settings. These are prepopulated with the same values as the primary cluster, but you can update them to meet specific requirements for that cluster.

- Port
- Number of replicas
- Subnet group
- Preferred Availability Zone(s)
- Security groups
- Customer Managed (Customer Master Key)
- Redis AUTH Token
- Enable automatic backups
- Backup retention period
- Backup window
- Maintenance window
- Topic for SNS notification

9. Choose Create. Doing this sets the status of the global datastore to Creating. The status transitions to Modifying after the primary cluster is associated to the global datastore and the secondary cluster is in Associating status.

After the primary cluster and secondary clusters are associated with the global datastore, the status changes to Available. At this point, you have a primary cluster that accepts reads and writes and secondary clusters that accept reads replicated from the primary cluster.

The Redis page is updated to indicate whether a cluster is part of a global datastore, including:

- **Global Datastore** – The name of the global datastore to which the cluster belongs.
- **Global Datastore Role** – The role of the cluster, either primary or secondary.

You can add up to one additional secondary cluster in a different AWS Region. For more information, see Adding a Region to a Global Datastore (p. 121).

**Creating a New Global Datastore Using a New Primary Cluster**

If you choose to create a new global datastore, use the following procedure.

**To create a new global datastore**

1. Sign in to the AWS Management Console and open the ElastiCache console at https://console.aws.amazon.com/elasticache/
2. On the navigation pane, choose Global Datastore and then choose Create.
3. Under Create Global Datastore, do the following:
   a. Enter a value for **Global Datastore Name suffix**. ElastiCache uses the suffix to generate a unique name for the global datastore. You can search for the global datastore by using the suffix that you specify here.
   b. (Optional) Enter a value for **Global Datastore Description**.
4. Under **Primary cluster details**, for **Region**, choose an available AWS Region and one of the following options:
   - Creating a New Regional Cluster as a Primary (p. 119).
   - Using an Existing Cluster as Primary Cluster (p. 120)

**Creating a New Regional Cluster as a Primary**

To create a new regional cluster to serve as the global datastore's primary cluster, do the following:

**To create a new regional cluster as a primary**

1. For **Name**, enter a name for the cluster.
2. (Optional) For **Description** (optional), add a description for the cluster.
3. Configure the remaining settings to meet your specific requirements. When you select a parameter group to set the engine configuration values, that parameter group is applied to all clusters in the global datastore. On the **Parameter Groups** page, the yes/no **Global** attribute indicates whether a parameter group is part of a global datastore.
4. Choose **Next**.
5. Under **Secondary cluster details**, select a different AWS Region where the cluster will be stored.
6. Under **Redis settings**, enter a **Name** and, optionally, a **Description** for the cluster.
7. The following fields are pre-populated to match the primary cluster configuration and cannot be changed:
   - Engine version
   - Instance type
   - Node type
   - Number of shards
   - Parameter group

   **Note**
   ElastiCache auto-generates a new parameter group from values of the provided parameter group and applies the new parameter group to the cluster. Use this new parameter group to modify parameters on a global datastore. Each auto-generated parameter group is associated with one and only one cluster and, therefore, only one global datastore.

   **Note**
   You do have the option to supply a different encryption key by selecting **Customer Managed Customer Master Key** and selecting the key.

   - Encryption at rest
   - Redis AUTH

The remaining secondary cluster settings are pre-populated with the same values as the primary cluster, but the following can be updated to meet specific requirements for that cluster.

   - Port
   - Number of replicas
   - Subnet group
   - Preferred Availability Zone(s)
   - Security groups
   - Customer Managed (Customer Master Key)
• Redis AUTH Token
• Enable automatic backups
• Backup retention period
• Backup window
• Maintenance window
• Topic for SNS notification

8. Choose Create. This sets the status of the global datastore to Creating. After the primary cluster and secondary clusters are associated with the global datastore, the status changes to Available. You have a primary cluster that accepts reads and writes and a secondary cluster that accepts reads replicated from the primary cluster.

The Redis page is also updated to indicate whether a cluster is part of a global datastore, including the following:

• Global Datastore – The name of the global datastore to which the cluster belongs.
• Global Datastore Role – Reflects the role of the cluster, either primary or secondary.

You can add up to one additional secondary cluster in a different AWS Region. For more information, see Adding a Region to a Global Datastore (p. 121).

Using an Existing Cluster as Primary Cluster

If you choose this option, follow the steps in Creating a Global Datastore Using an Existing Cluster (p. 117) beginning with step 5.

Viewing Global Datastore Details

You can view the details of existing global datastores and also modify them on the Global Datastore page.

To view global datastore details

2. On the navigation pane, choose Global Datastore and then choose an available global datastore.

You can then examine the following global datastore properties:

• Global Datastore Name: The name of the global datastore
• Description: A description of the global datastore
• Status: Options include:
  • Creating
  • Modifying
  • Available
  • Deleting
  • Primary-Only - This status indicates the global datastore contains only a primary cluster. Either all secondary clusters are deleted or not successfully created.
• Cluster Mode: Either enabled or disabled
• Redis Engine Version: The Redis engine version running the global datastore
• Instance Node Type: The node type used for the global datastore
• Encryption at-rest: Either enabled or disabled
• **Encryption in-transit:** Either enabled or disabled
• **Redis AUTH:** Either enabled or disabled

You can make the following changes to the global datastore:

- **Adding a Region to a Global Datastore (p. 121)**
- **Removing a Region from a Global Datastore (p. 123)**
- **Promoting the Secondary Cluster to Primary (p. 123)**
- **Modifying a Global Datastore (p. 122)**

The Global Datastore page also lists the individual clusters that make up the global datastore and the following properties for each:

- **Region** - The AWS Region where the cluster is stored
- **Role** - Either primary or secondary
- **Cluster name** - The name of the cluster
- **Status** - Options include:
  - **Associating** - The cluster is in the process of being associated to the global datastore
  - **Associated** - The cluster is associated to the global datastore
  - **Disassociating** - The process of removing a secondary cluster from the global datastore using the global datastore name. After this, the secondary cluster no longer receives updates from the primary cluster but it remains as a standalone cluster in that AWS Region.
  - **Disassociated** - The secondary cluster has been removed from the global datastore and is now a standalone cluster in its AWS Region.
- **Global Datastore Replica lag** – Shows one value per secondary AWS Region in the global datastore. This is the lag between the secondary Region's master node and the primary region's master node. For cluster mode enabled Redis, the lag indicates the maximum delay among the shards.

### Adding a Region to a Global Datastore

You can add up to one additional AWS Region to an existing global datastore. In this scenario, you are creating a read-only cluster in a separate AWS Region that receives automatic and asynchronous updates from the primary cluster.

**To add an AWS Region to a global datastore**

2. On the navigation pane, choose Global Datastore and then select a global datastore under **Global Datastore Name**.
3. Choose **Add Region**.
4. Choose the AWS Region where the secondary cluster is to reside.
5. Under **Redis settings**, enter a **Name** and, optionally, a **Description** for the cluster.
6. Keep the following options as they are. They're prepopulated to match the primary cluster configuration, you can't change them.
   - Engine version
   - Instance type
   - Node type
   - Number of shards
• Parameter group

  **Note**
  ElastiCache auto-generates a new parameter group from values of the provided parameter group and applies the new parameter group to the cluster. Use this new parameter group to modify parameters on a global datastore. Each auto-generated parameter group is associated with one and only one cluster and, therefore, only one global datastore.

• Encryption at rest

  **Note**
  You do have the option to supply a different encryption key by selecting **Customer Managed Customer Master Key** and selecting the key.

• Encryption in-transit

7. (Optional) As needed, update the remaining secondary cluster settings. These are prepopulated with the same values as the primary cluster, but you can update them to meet specific requirements for that cluster.

• Port

• Number of replicas

• Subnet group

• Preferred Availability Zone(s)

• Security groups

• Customer Managed (Customer Master Key)

• Redis AUTH Token

• Enable automatic backups

• Backup retention period

• Backup window

• Maintenance window

• Topic for SNS notification

8. Choose *Add*.

## Modifying a Global Datastore

You can modify properties of regional clusters. Only one modify operation can be in progress on a global datastore, with the exception of promoting a secondary cluster to primary. For more information, see **Promoting the Secondary Cluster to Primary** (p. 123).

**To modify a global datastore**


2. On the navigation pane, choose **Global Datastore** and then for **Global Datastore Name**, choose a global datastore.

3. Choose **Modify** and choose among the following options:

   • **Modify description** – Update the description of the global datastore

   • **Modify engine version** – Only Redis engine version 5.0.6 or later is available.

   • **Modify node type** – Scale regional clusters both vertically (scaling up and down) and horizontally (scaling in and out). Options include the R5 and M5 node families. For more information on node types, see **Supported Node Types** (p. 46).
• **Modify Automatic Failover** – Enable or disable Automatic Failover. When you enable failover and primary nodes in regional clusters die, ElastiCache then fails over to one of the regional replicas. For more information, see Auto Failover.

For Redis clusters with cluster-mode enabled:

• **Add shards** – Enter the number of shards to add and optionally specify one or more Availability Zones.

• **Delete shards** – Choose shards to be deleted in each AWS Region.

• **Rebalance shards** – Rebalance the slot distribution to ensure uniform distribution across existing shards in the cluster.

To modify a global datastore's parameters, modify the parameter group of any member cluster for the global datastore. ElastiCache applies this change to all clusters within that global datastore automatically. To modify the parameter group of that cluster, use the Redis console or the ModifyCacheCluster API. For more information, see Modifying a Parameter Group (p. 146). When you modify the parameter group of any cluster contained within a global datastore, it is applied to all the clusters within that global datastore.

To reset an entire parameter group or specific parameters, use the ResetCacheParameterGroup API.

**Promoting the Secondary Cluster to Primary**

If the primary cluster or AWS Region becomes unavailable or is experiencing performance issues, you can promote a secondary cluster to primary. Promotion is allowed anytime, even if other modifications are in progress. You can also issue multiple promotions in parallel and the global datastore resolves to one primary eventually. If you promote multiple secondary clusters simultaneously, ElastiCache for Redis doesn't guarantee which one ultimately resolves to primary.

**To promote a secondary cluster to primary**

2. On the navigation pane, choose Global Datastore under Redis.
3. Select the global datastore name to view the details
4. Choose the Secondary cluster
5. Choose Promote to primary.

You're then prompted to confirm your decision with the following warning: Promoting a region to primary will make the cluster in this region as read/writable. Are you sure you want to promote the secondary cluster to primary?

The current primary cluster in primary region will become secondary and will stop accepting writes after this operation completes. Please ensure you update your application stack to direct traffic to the new primary region.

6. Choose Confirm if you want to continue the promotion or Cancel if you don't.

If you choose to confirm, your global datastore moves to a Modifying state and is unavailable until the promotion is complete.

**Removing a Region from a Global Datastore**

You can remove an AWS Region from a global datastore by using the following procedure.
To remove an AWS Region from a global datastore

2. On the navigation pane, choose Global Datastore under Redis.
3. Select a global datastore.
4. Choose the Region you want to remove.
5. Choose Remove region.

   Note
   This option is only available for secondary clusters.

   You're then be prompted to confirm your decision with the following warning: Removing the region will remove your only available cross region replica for the primary cluster. Your primary cluster will no longer be set up for disaster recovery and improved read latency in remote region. Are you sure you want to remove the selected region from the global datastore?
6. Choose Confirm if you want to continue the promotion or Cancel if you don't.

If you choose confirm, the AWS Region is removed and the secondary cluster no longer receives replication updates.

Deleting a Global Datastore

To delete a global datastore, first remove all secondary clusters. For more information, see Removing a Region from a Global Datastore (p. 123). Doing this leaves the global datastore in primary-only status.

To delete a global datastore

2. On the navigation pane, choose Global Datastore under Redis.
3. Under Global Datastore Name choose the global datastore you want to delete and then choose Delete.

   You're then be prompted to confirm your decision with the following warning: Are you sure you want to delete this Global Datastore?
4. Choose Delete.

The global datastore transitions to Deleting status.

Using Global Datastores (CLI)

You can use the AWS Command Line Interface (AWS CLI) to control multiple AWS services from the command line and automate them through scripts. You can use the AWS CLI for ad hoc (one-time) operations.

Downloading and Configuring the AWS CLI

The AWS CLI runs on Windows, macOS, or Linux. Use the following procedure to download and configure it.

To download, install, and configure the CLI

1. Download the AWS CLI on the AWS Command Line Interface webpage.
2. Follow the instructions for Installing the AWS CLI and Configuring the AWS CLI in the AWS Command Line Interface User Guide.

Using the AWS CLI with Global Datastores

Use the following CLI operations to work with global datastores:

- **create-global-replication-group**

  ```bash
  aws elasticache create-global-replication-group
  --global-replication-group-id-suffix my global datastore
  --primary-replication-group-id sample-repl-group
  --global-replication-group-description an optional description of the global datastore
  
  --automatic-failover-enabled yes/no
  --cache-node-type node type
  --engine-version engine version
  --apply-immediately
  --global-replication-group-description description
  ```

- **create-replication-group** – Use this operation to create secondary clusters for a Global Datastore by supplying the name of the Global Datastore to the `--global-replication-group-id` parameter.

  ```bash
  aws elasticache create-replication-group
  --replication-group-id secondary replication group name
  --replication-group-description "Replication group description"
  --global-replication-group-id global datastore name
  ```

- **describe-global-replication-groups**

  ```bash
  aws elasticache describe-global-replication-groups
  --global-replication-group-id my global datastore
  --show-member-info an optional parameter that returns a list of the primary and secondary clusters that make up the global datastore
  ```

- **modify-global-replication-group**

  ```bash
  aws elasticache modify-global-replication-group
  --global-replication-group-id my global datastore
  --automatic-failover-enabled yes/no
  --cache-node-type node type
  --engine-version engine version
  --apply-immediately
  --global-replication-group-description description
  ```

- **delete-global-replication-group**

  ```bash
  aws elasticache delete-global-replication-group
  --global-replication-group-id my global datastore
  --retain-primary-replication-group defaults to true
  ```

- **disassociate-global-replication-group**

  ```bash
  aws elasticache disassociate-global-replication-group
  --global-replication-group-id my Global Datastore
  --replication-group-id my secondary cluster
  --replication-group-region the AWS Region in which the secondary cluster resides
  ```

- **failover-global-replication-group**

  ```bash
  aws elasticache failover-replication-group
  ````
Finding Connection Endpoints

Your application connects to your cluster using endpoints. An endpoint is a node or cluster's unique address.

Which endpoints to use

- Memcached cluster, If you use Automatic Discovery, you can use the cluster's configuration endpoint to configure your Memcached client. This means you must use a client that supports Automatic Discovery.

  If you don't use Automatic Discovery, you must configure your client to use the individual node endpoints for reads and writes. You must also keep track of them as you add and remove nodes.

The following sections guide you through discovering the endpoints you'll need for the engine you're running.

Topics

- Finding a Cluster's Endpoints (Console) (p. 128)
- Finding Endpoints (AWS CLI) (p. 130)
• Finding Endpoints (ElastiCache API) (p. 132)
Finding a Cluster's Endpoints (Console)

All Memcached endpoints are read/write endpoints. To connect to nodes in a Memcached cluster your application can use either the endpoints for each node, or the cluster's configuration endpoint along with Automatic Discovery. To use Automatic Discovery you must use a client that supports Automatic Discovery.

When using Automatic Discovery, your client application connects to your Memcached cluster using the configuration endpoint. As you scale your cluster by adding or removing nodes, your application will automatically "know" all the nodes in the cluster and be able to connect to any of them. Without Automatic Discovery your application would have to do this, or you'd have to manually update endpoints in your application each time you added or removed a node. For additional information on Automatic Discovery, see Automatically Identify Nodes in your Memcached Cluster (p. 55).

The following procedure demonstrates how to find and copy a cluster's configuration endpoint or any of the node endpoints using the ElastiCache console.

**To find and copy the endpoints for a Memcached cluster (console)**

2. From the navigation pane, choose **Memcached**.
   The cache clusters screen will appear with a list of Memcached clusters.
3. Find the Memcached cluster you want the endpoints for.
   If all you want is the configuration endpoint, you're done. The configuration endpoint is in the **Configuration Endpoint** column and looks something like this, `clusterName.xxxxxx.cfg.usw2.cache.amazonaws.com:port`.
   If you want to also see the individual node endpoints or copy any of the endpoints to your clipboard, choose **Copy Node Endpoint**.

   **Endpoints for a Memcached cluster**

4. To copy an endpoint to your clipboard:
   a. On the **Copy Node Endpoint** screen, highlight the endpoint you want to copy.
   b. Right-click the highlighted endpoint, and then choose **Copy** from the context menu.

   The highlighted endpoint is now copied to your clipboard.

Configuration and node endpoints look very similar. The differences are highlighted with **bold** following.
Important
If you choose to create a CNAME for your Memcached configuration endpoint, in order for your automatic discovery client to recognize the CNAME as a configuration endpoint, you must include `.cfg.` in the CNAME.

```
myclusternamexxxxxx.cfg.usw2.cache.amazonaws.com:port  # configuration endpoint contains "cfg"
myclusternamexxxxxx.0001.usw2.cache.amazonaws.com:port  # node endpoint for node 0001
```
Finding Endpoints (AWS CLI)

You can use the AWS CLI for Amazon ElastiCache to discover the endpoints for nodes and clusters.

Topics
- Finding Endpoints for Nodes and Clusters (AWS CLI) (p. 130)

Finding Endpoints for Nodes and Clusters (AWS CLI)

You can use the AWS CLI to discover the endpoints for a cluster and its nodes with the `describe-cache-clusters` command. For Memcached clusters, the command returns the configuration endpoint. If you include the optional parameter `--show-cache-node-info`, the command will also return the endpoints of the individual nodes in the cluster.

Example

The following command retrieves the configuration endpoint (`ConfigurationEndpoint`) and individual node endpoints (`Endpoint`) for the Memcached cluster `mycluster`.

For Linux, macOS, or Unix:

```
aws elasticache describe-cache-clusters \
  --cache-cluster-id mycluster  \
  --show-cache-node-info
```

For Windows:

```
aws elasticache describe-cache-clusters ^
  --cache-cluster-id mycluster ^
  --show-cache-node-info
```

Output from the above operation should look something like this (JSON format).

```json
{
  "CacheClusters": [
  {
    "Engine": "memcached",
    "CacheNodes": [
      {
        "CacheNodeId": "0001",
        "Endpoint": {
          "Port": 11211,
          "Address": "mycluster.1abc4d.0001.usw2.cache.amazonaws.com"
        },
        "CacheNodeStatus": "available",
        "ParameterGroupStatus": "in-sync",
        "CacheNodeCreateTime": "2016-09-22T21:30:29.967Z",
        "CustomerAvailabilityZone": "us-west-2b"
      },
      {
        "CacheNodeId": "0002",
        "Endpoint": {
          "Port": 11211,
          "Address": "mycluster.1abc4d.0002.usw2.cache.amazonaws.com"
        },
        "CacheNodeStatus": "available",
        "ParameterGroupStatus": "in-sync",
        "CacheNodeCreateTime": "2016-09-22T21:30:29.967Z",
        "CustomerAvailabilityZone": "us-west-2b"
      }
    ]
  }]
```
Important

If you choose to create a CNAME for your Memcached configuration endpoint, in order for your PHP client to recognize the CNAME as a configuration endpoint, you must include .cfg. in the CNAME. For example, mycluster.cfg.local in your php.ini file for the session.save_path parameter.

For more information, see the topic describe-cache-clusters.
Finding Endpoints (ElastiCache API)

You can use the Amazon ElastiCache API to discover the endpoints for nodes and clusters.

Topics
- Finding Endpoints for Nodes and Clusters (ElastiCache API) (p. 132)

Finding Endpoints for Nodes and Clusters (ElastiCache API)

You can use the ElastiCache API to discover the endpoints for a cluster and its nodes with the DescribeCacheClusters action. For Memcached clusters, the command returns the configuration endpoint. If you include the optional parameter ShowCacheNodeInfo, the action also returns the endpoints of the individual nodes in the cluster.

Example

The following command retrieves the configuration endpoint (ConfigurationEndpoint) and individual node endpoints (Endpoint) for the Memcached cluster mycluster.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheClusters
&CacheClusterId=mycluster
&ShowCacheNodeInfo=true
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&Version=2015-02-02
&X-Amz-Credential=<credential>
```

Important

If you choose to create a CNAME for your Memcached configuration endpoint, in order for your PHP client to recognize the CNAME as a configuration endpoint, you must include .cfg. in the CNAME. For example, mycluster.cfg.local in your php.ini file for the session.save_path parameter.

Scaling ElastiCache for Memcached Clusters

The amount of data your application needs to process is seldom static. It increases and decreases as your business grows or experiences normal fluctuations in demand. If you self-manage your cache, you need to provision sufficient hardware for your demand peaks, which can be expensive. By using Amazon ElastiCache you can scale to meet current demand, paying only for what you use. ElastiCache enables you to scale your cache to match demand.

The following helps you find the correct topic for the scaling actions you want to perform.

Scaling Memcached Clusters

<table>
<thead>
<tr>
<th>Action</th>
<th>Topic/Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling out</td>
<td>Adding Nodes to a Cluster (p. 94)</td>
</tr>
<tr>
<td>Scaling in</td>
<td>Removing Nodes from a Cluster (p. 99)</td>
</tr>
<tr>
<td>Changing node types</td>
<td>Scaling Memcached Vertically (p. 133)</td>
</tr>
</tbody>
</table>
Memcached clusters are composed of 1 to 20 nodes. Scaling a Memcached cluster out and in is as easy as adding or removing nodes from the cluster.

If you need more than 20 nodes in a Memcached cluster, or more than 100 nodes total in an AWS Region, fill out the ElastiCache Limit Increase Request form at https://aws.amazon.com/contact-us/elasticache-node-limit-request/.

Because you can partition your data across all the nodes in a Memcached cluster, scaling up to a node type with greater memory is seldom required. However, because the Memcached engine does not persist data, if you do scale to a different node type, your new cluster starts out empty unless your application populates it.

Topics
- Scaling Memcached Horizontally (p. 133)
- Scaling Memcached Vertically (p. 133)

Scaling Memcached Horizontally

The Memcached engine supports partitioning your data across multiple nodes. Because of this, Memcached clusters scale horizontally easily. A Memcached cluster can have from 1 to 20 nodes. To horizontally scale your Memcached cluster, merely add or remove nodes.

If you need more than 20 nodes in a Memcached cluster, or more than 100 nodes total in an AWS Region, fill out the ElastiCache Limit Increase Request form at https://aws.amazon.com/contact-us/elasticache-node-limit-request/.

The following topics detail how to scale your Memcached cluster out or in by adding or removing nodes.
- Adding Nodes to a Cluster (p. 94)
- Removing Nodes from a Cluster (p. 99)

Each time you change the number of nodes in your Memcached cluster, you must re-map at least some of your keyspace so it maps to the correct node. For more detailed information on load balancing your Memcached cluster, see Configuring Your ElastiCache Client for Efficient Load Balancing (p. 31).

If you use auto discovery on your Memcached cluster, you do not need to change the endpoints in your application as you add or remove nodes. For more information on auto discovery, see Automatically Identify Nodes in your Memcached Cluster (p. 55). If you do not use auto discovery, each time you change the number of nodes in your Memcached cluster you must update the endpoints in your application.

Scaling Memcached Vertically

When you scale your Memcached cluster up or down, you must create a new cluster. Memcached clusters always start out empty unless your application populates it.

Important
- If you are scaling down to a smaller node type, be sure that the smaller node type is adequate for your data and overhead. For more information, see Choosing Your Memcached Node Size (p. 80).

Topics
- Scaling Memcached Vertically (Console) (p. 134)
- Scaling Memcached Vertically (AWS CLI) (p. 134)
- Scaling Memcached Vertically (ElastiCache API) (p. 134)
Scaling Memcached Vertically (Console)

The following procedure walks you through scaling your cluster vertically using the ElastiCache console.

To scale a Memcached cluster vertically (console)

1. Create a new cluster with the new node type. For more information, see Creating a Memcached Cluster (Console) (p. 82).
2. In your application, update the endpoints to the new cluster's endpoints. For more information, see Finding a Cluster's Endpoints (Console) (p. 128).
3. Delete the old cluster. For more information, see Using the AWS Management Console (p. 105).

Scaling Memcached Vertically (AWS CLI)

The following procedure walks you through scaling your Memcached cache cluster vertically using the AWS CLI.

To scale a Memcached cache cluster vertically (AWS CLI)

1. Create a new cache cluster with the new node type. For more information, see Creating a Cluster (AWS CLI) (p. 84).
2. In your application, update the endpoints to the new cluster's endpoints. For more information, see Finding Endpoints (AWS CLI) (p. 130).
3. Delete the old cache cluster. For more information, see Using the AWS CLI (p. 105).

Scaling Memcached Vertically (ElastiCache API)

The following procedure walks you through scaling your Memcached cache cluster vertically using the ElastiCache API.

To scale a Memcached cache cluster vertically (ElastiCache API)

1. Create a new cache cluster with the new node type. For more information, see Creating a Cluster (ElastiCache API) (p. 85).
2. In your application, update the endpoints to the new cache cluster's endpoints. For more information, see Finding Endpoints (ElastiCache API) (p. 132).
3. Delete the old cache cluster. For more information, see Using the ElastiCache API (p. 106).
Configuring Engine Parameters Using Parameter Groups

Amazon ElastiCache uses parameters to control the runtime properties of your nodes and clusters. Generally, newer engine versions include additional parameters to support the newer functionality. For tables of parameters, see Memcached Specific Parameters (p. 151).

As you would expect, some parameter values, such as `max_cache_memory`, are determined by the engine and node type. For a table of these parameter values by node type, see Memcached Node-Type Specific Parameters (p. 158).

**Note**
For a list of Memcached-specific parameters, see Memcached Specific Parameters.

**Topics**
- Parameter Management (p. 136)
- Cache Parameter Group Tiers (p. 137)
- Creating a Parameter Group (p. 137)
- Listing Parameter Groups by Name (p. 141)
- Listing a Parameter Group's Values (p. 145)
- Modifying a Parameter Group (p. 146)
- Deleting a Parameter Group (p. 149)
- Memcached Specific Parameters (p. 151)
Parameter Management

Parameters are grouped together into named parameter groups for easier parameter management. A parameter group represents a combination of specific values for the parameters that are passed to the engine software during startup. These values determine how the engine processes on each node will behave at runtime. The parameter values on a specific parameter group apply to all nodes that are associated with the group, regardless of which cluster they belong to.

To fine-tune your cluster's performance, you can modify some parameter values or change the cluster's parameter group.

- You cannot modify or delete the default parameter groups. If you need custom parameter values, you must create a custom parameter group.
- The parameter group family and the cluster you're assigning it to must be compatible. For example, if your cluster is running Memcached version 1.4.8, you can only use parameter groups, default or custom, from the Memcached 1.4 family.
- If you change a cluster's parameter group, the values for any conditionally modifiable parameter must be the same in both the current and new parameter groups.
- When you change a cluster's parameters, the change is applied to the cluster either immediately or after the cluster is restarted. This is true whether you change the cluster's parameter group itself or a parameter value within the cluster's parameter group. To determine when a particular parameter change is applied, see the Changes Take Effect column in the tables for Memcached Specific Parameters (p. 151). For information on rebooting a cluster, see Rebooting a Cluster (p. 93).
Cache Parameter Group Tiers

Amazon ElastiCache has three tiers of cache parameter groups as shown following.

Global Default

The top-level root parameter group for all Amazon ElastiCache customers in the region.

The global default cache parameter group:
- Is reserved for ElastiCache and not available to the customer.

Customer Default

A copy of the Global Default cache parameter group which is created for the customer's use.

The Customer Default cache parameter group:
- Is created and owned by ElastiCache.
- Is available to the customer for use as a cache parameter group for any clusters running an engine version supported by this cache parameter group.
- Cannot be edited by the customer.

Customer Owned

A copy of the Customer Default cache parameter group. A Customer Owned cache parameter group is created whenever the customer creates a cache parameter group.

The Customer Owned cache parameter group:
- Is created and owned by the customer.
- Can be assigned to any of the customer's compatible clusters.
- Can be modified by the customer to create a custom cache parameter group.

Not all parameter values can be modified. For more information, see Memcached Specific Parameters (p. 151).

Creating a Parameter Group

You need to create a new parameter group if there is one or more parameter values that you want changed from the default values. You can create a parameter group using the ElastiCache console, the AWS CLI, or the ElastiCache API.
Creating a Parameter Group (Console)

The following procedure shows how to create a parameter group using the ElastiCache console.

To create a parameter group using the ElastiCache console

2. To see a list of all available parameter groups, in the left hand navigation pane choose Parameter Groups.
3. To create a parameter group, choose Create Parameter Group.
   
   The Create Parameter Group screen will appear.
4. From the Family list, choose the parameter group family that will be the template for your parameter group.

   The parameter group family, such as memcached1.4, defines the actual parameters in your parameter group and their initial values. The parameter group family must coincide with the cluster's engine and version.
5. In the Name box, type in a unique name for this parameter group.

   When creating a cluster or modifying a cluster's parameter group, you will choose the parameter group by its name. Therefore, we recommend that the name be informative and somehow identify the parameter group's family.

   Parameter group naming constraints are as follows:
   • Must begin with an ASCII letter.
   • Can only contain ASCII letters, digits, and hyphens.
   • Must be 1–255 characters long.
   • Can't contain two consecutive hyphens.
   • Can't end with a hyphen.
6. In the Description box, type in a description for the parameter group.
7. To create the parameter group, choose Create.

   To terminate the process without creating the parameter group, choose Cancel.
8. When the parameter group is created, it will have the family's default values. To change the default values you must modify the parameter group. For more information, see Modifying a Parameter Group (p. 146).

Creating a Parameter Group (AWS CLI)

To create a parameter group using the AWS CLI, use the command create-cache-parameter-group with these parameters.

• --cache-parameter-group-name — The name of the parameter group.

   Parameter group naming constraints are as follows:
   • Must begin with an ASCII letter.
   • Can only contain ASCII letters, digits, and hyphens.
   • Must be 1–255 characters long.
   • Can't contain two consecutive hyphens.
   • Can't end with a hyphen.
Creating a Parameter Group

- `--cache-parameter-group-family` — The engine and version family for the parameter group.
- `--description` — A user supplied description for the parameter group.

**Example**

The following example creates a parameter group named `myMem14` using the `memcached1.4` family as the template.

For Linux, macOS, or Unix:

```bash
aws elasticache create-cache-parameter-group \
  --cache-parameter-group-name myMem14 \
  --cache-parameter-group-family memcached1.4 \
  --description "My first parameter group"
```

For Windows:

```bash
aws elasticache create-cache-parameter-group ^
  --cache-parameter-group-name myMem14 ^
  --cache-parameter-group-family memcached1.4 ^
  --description "My first parameter group"
```

The output from this command should look something like this.

```json
{
  "CacheParameterGroup": {
    "CacheParameterGroupName": "myMem14",
    "CacheParameterGroupFamily": "memcached1.4",
    "Description": "My first parameter group"
  }
}
```

When the parameter group is created, it will have the family's default values. To change the default values you must modify the parameter group. For more information, see Modifying a Parameter Group (p. 146).

For more information, see `create-cache-parameter-group`.

**Creating a Parameter Group (ElastiCache API)**

To create a parameter group using the ElastiCache API, use the `CreateCacheParameterGroup` action with these parameters.

- `ParameterGroupName` — The name of the parameter group.

  Parameter group naming constraints are as follows:
  - Must begin with an ASCII letter.
  - Can only contain ASCII letters, digits, and hyphens.
  - Must be 1–255 characters long.
  - Can't contain two consecutive hyphens.
  - Can't end with a hyphen.

- `CacheParameterGroupFamily` — The engine and version family for the parameter group. For example, `memcached1.4`.
- `Description` — A user supplied description for the parameter group.

API Version 2015-02-02

139
Example

The following example creates a parameter group named *myMem14* using the memcached1.4 family as the template.

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateCacheParameterGroup
  &CacheParameterGroupFamily=memcached1.4
  &CacheParameterGroupName=myMem14
  &Description=My%20first%20parameter%20group
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &Version=2015-02-02
  &X-Amz-Credential=<credential>
```

The response from this action should look something like this.

```
  <CreateCacheParameterGroupResult>
    <CacheParameterGroup>
      <CacheParameterGroupName>myMem14</CacheParameterGroupName>
      <CacheParameterGroupFamily>memcached1.4</CacheParameterGroupFamily>
      <Description>My first parameter group</Description>
    </CacheParameterGroup>
  </CreateCacheParameterGroupResult>
  <ResponseMetadata>
    <RequestId>d8465952-af48-11e0-8d36-859edca6f4b8</RequestId>
  </ResponseMetadata>
</CreateCacheParameterGroupResponse>
```

When the parameter group is created, it will have the family's default values. To change the default values you must modify the parameter group. For more information, see Modifying a Parameter Group (p. 146).

For more information, see CreateCacheParameterGroup.
Listing Parameter Groups by Name

You can list the parameter groups using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Listing Parameter Groups by Name (Console)

The following procedure shows how to view a list of the parameter groups using the ElastiCache console.

To list parameter groups using the ElastiCache console

2. To see a list of all available parameter groups, in the left hand navigation pane choose Parameter Groups.

Listing Parameter Groups by Name (AWS CLI)

To generate a list of parameter groups using the AWS CLI, use the command describe-cache-parameter-groups. If you provide a parameter group’s name, only that parameter group will be listed. If you do not provide a parameter group’s name, up to --max-records parameter groups will be listed. In either case, the parameter group’s name, family, and description are listed.

Example

The following sample code lists the parameter group myMem14.

For Linux, macOS, or Unix:

```
aws elasticache describe-cache-parameter-groups \\
--cache-parameter-group-name myMem14
```

For Windows:

```
aws elasticache describe-cache-parameter-groups ^
--cache-parameter-group-name myMem14
```

The output of this command will look something like this, listing the name, family, and description for the parameter group.

```
{
  "CacheParameterGroups": [
  {
    "CacheParameterGroupName": "myMem14",
    "CacheParameterGroupFamily": "memcached1.4",
    "Description": "My first parameter group"
  }
  ]
}
```

Example

The following sample code lists up to 10 parameter groups.

```
aws elasticache describe-cache-parameter-groups --max-records 20
```
The JSON output of this command will look something like this, listing the name, family, description and, in the case of redis5.6 whether the parameter group is part of a Global Datstore (isGlobal), for each parameter group.

```
{
    "CacheParameterGroups": [
        {
            "CacheParameterGroupName": "custom-redis32",
            "CacheParameterGroupFamily": "redis3.2",
            "Description": "custom parameter group with reserved-memory > 0"
        },
        {
            "CacheParameterGroupName": "default.memcached1.4",
            "CacheParameterGroupFamily": "memcached1.4",
            "Description": "Default parameter group for memcached1.4"
        },
        {
            "CacheParameterGroupName": "default.redis2.6",
            "CacheParameterGroupFamily": "redis2.6",
            "Description": "Default parameter group for redis2.6"
        },
        {
            "CacheParameterGroupName": "default.redis2.8",
            "CacheParameterGroupFamily": "redis2.8",
            "Description": "Default parameter group for redis2.8"
        },
        {
            "CacheParameterGroupName": "default.redis3.2",
            "CacheParameterGroupFamily": "redis3.2",
            "Description": "Default parameter group for redis3.2"
        },
        {
            "CacheParameterGroupName": "default.redis3.2.cluster.on",
            "CacheParameterGroupFamily": "redis3.2",
            "Description": "Customized default parameter group for redis3.2 with cluster mode on"
        },
        {
            "CacheParameterGroupName": "default.redis5.6.cluster.on",
            "CacheParameterGroupFamily": "redis5.0",
            "Description": "Customized default parameter group for redis5.6 with cluster mode on",
            "isGlobal": "yes"
        }
    ]
}
```

For more information, see [describe-cache-parameter-groups](https://elasticache.us-west-2.amazonaws.com/)

### Listing Parameter Groups by Name (ElastiCache API)

To generate a list of parameter groups using the ElastiCache API, use the DescribeCacheParameterGroups action. If you provide a parameter group's name, only that parameter group will be listed. If you do not provide a parameter group's name, up to MaxRecords parameter groups will be listed. In either case, the parameter group's name, family, and description are listed.

**Example**

The following sample code lists the parameter group *myMem14*.

https://elasticache.us-west-2.amazonaws.com/
The response from this action will look something like this, listing the name, family, and description for each parameter group.

```xml
  <DescribeCacheParameterGroupsResult>
    <CacheParameterGroups>
      <CacheParameterGroup>
        <CacheParameterGroupName>myMem14</CacheParameterGroupName>
        <CacheParameterGroupFamily>memcached1.4</CacheParameterGroupFamily>
        <Description>My custom Memcached 1.4 parameter group</Description>
      </CacheParameterGroup>
    </CacheParameterGroups>
  </DescribeCacheParameterGroupsResult>
  <ResponseMetadata>
    <RequestId>3540cc3d-af48-11e0-97f9-279771c4477e</RequestId>
  </ResponseMetadata>
</DescribeCacheParameterGroupsResponse>
```

**Example**

The following sample code lists up to 10 parameter groups.

```xml
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheParameterGroups
&MaxRecords=10
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&Version=2015-02-02
&X-Amz-Credential=<credential>
```

The response from this action will look something like this, listing the name, family, description and, in the case of redis5.6 if the parameter group belongs to a Global Datastore (isGlobal), for each parameter group.

```xml
  <DescribeCacheParameterGroupsResult>
    <CacheParameterGroups>
      <CacheParameterGroup>
        <CacheParameterGroupName>myRedis28</CacheParameterGroupName>
        <CacheParameterGroupFamily>redis2.8</CacheParameterGroupFamily>
        <Description>My custom Redis 2.8 parameter group</Description>
      </CacheParameterGroup>
      <CacheParameterGroup>
        <CacheParameterGroupName>myMem14</CacheParameterGroupName>
        <CacheParameterGroupFamily>memcached1.4</CacheParameterGroupFamily>
        <Description>My custom Memcached 1.4 parameter group</Description>
      </CacheParameterGroup>
      <CacheParameterGroup>
        <CacheParameterGroupName>myRedis56</CacheParameterGroupName>
        <CacheParameterGroupFamily>redis5.0</CacheParameterGroupFamily>
        <Description>My custom Redis 5.0 parameter group</Description>
      </CacheParameterGroup>
    </CacheParameterGroups>
  </DescribeCacheParameterGroupsResult>
  <ResponseMetadata>
    <RequestId>3540cc3d-af48-11e0-97f9-279771c4477e</RequestId>
  </ResponseMetadata>
</DescribeCacheParameterGroupsResponse>
```
Example

The following sample code lists up to 10 parameter groups.

```
https://elastcache.us-west-2.amazonaws.com/
  ?Action=DescribeCacheParameterGroups
  &MaxRecords=10
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &Version=2015-02-02
  &X-Amz-Credential=<credential>
```

The response from this action will look something like this, listing the name, family, description and, in the case of `redis5.6`, whether the parameter group is part of a Global Datastore (isGlobal), for each parameter group.

```
  <DescribeCacheParameterGroupsResult>
    <CacheParameterGroups>
      <CacheParameterGroup>
        <CacheParameterGroupName>myRedis28</CacheParameterGroupName>
        <CacheParameterGroupFamily>redis2.8</CacheParameterGroupFamily>
        <Description>My custom Redis 2.8 parameter group</Description>
      </CacheParameterGroup>
      <CacheParameterGroup>
        <CacheParameterGroupName>myMem14</CacheParameterGroupName>
        <CacheParameterGroupFamily>memcached1.4</CacheParameterGroupFamily>
        <Description>My custom Memcached 1.4 parameter group</Description>
      </CacheParameterGroup>
      <CacheParameterGroup>
        <CacheParameterGroupName>myRedis56</CacheParameterGroupName>
        <CacheParameterGroupFamily>redis5.0</CacheParameterGroupFamily>
        <Description>My custom Redis 5.0 parameter group</Description>
      </CacheParameterGroup>
    </CacheParameterGroups>
    <CacheParameterGroup>
      <CacheParameterGroupName>myRedis56</CacheParameterGroupName>
      <CacheParameterGroupFamily>redis5.6</CacheParameterGroupFamily>
      <Description>My custom Redis 5.6 parameter group</Description>
      <isGlobal>yes</isGlobal>
    </CacheParameterGroup>
  </DescribeCacheParameterGroupsResult>
  <ResponseMetadata>
    <RequestId>3540cc3d-af48-11e0-97f9-279771c4477e</RequestId>
  </ResponseMetadata>
</DescribeCacheParameterGroupsResponse>
```

For more information, see `DescribeCacheParameterGroups`. 
Listing a Parameter Group's Values

You can list the parameters and their values for a parameter group using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Listing a Parameter Group's Values (Console)

The following procedure shows how to list the parameters and their values for a parameter group using the ElastiCache console.

To list a parameter group's parameters and their values using the ElastiCache console

2. To see a list of all available parameter groups, in the left hand navigation pane choose Parameter Groups.
3. Choose the parameter group for which you want to list the parameters and values by choosing the box to the left of the parameter group's name.

   The parameters and their values will be listed at the bottom of the screen. Due to the number of parameters, you may have to scroll up and down to find the parameter you're interested in.

Listing a Parameter Group's Values (AWS CLI)

To list a parameter group's parameters and their values using the AWS CLI, use the command describe-cache-parameters.

Example

The following sample code list all the parameters and their values for the parameter group myMem14.

For Linux, macOS, or Unix:

```bash
aws elasticache describe-cache-parameters \
--cache-parameter-group-name myMem14
```

For Windows:

```bash
aws elasticache describe-cache-parameters ^
--cache-parameter-group-name myMem14
```

For more information, see describe-cache-parameters.

Listing a Parameter Group's Values (ElastiCache API)

To list a parameter group's parameters and their values using the ElastiCache API, use the DescribeCacheParameters action.

Example

The following sample code list all the parameters for the parameter group myMem14.

```bash
https://elasticache.us-west-2.amazonaws.com/api
```

API Version 2015-02-02
145
Modifying a Parameter Group

Important
You cannot modify any default parameter group.

You can modify some parameter values in a parameter group. These parameter values are applied to clusters associated with the parameter group. For more information on when a parameter value change is applied to a parameter group, see Memcached Specific Parameters (p. 151).

Modifying a Parameter Group (Console)

The following procedure shows how to change the binding_protocol parameter’s value using the ElastiCache console. You would use the same procedure to change the value of any parameter.
To change a parameter's value using the ElastiCache console

2. To see a list of all available parameter groups, in the left hand navigation pane choose Parameter Groups.
3. Choose the parameter group you want to modify by choosing the box to the left of the parameter group's name.
   The parameter group's parameters will be listed at the bottom of the screen. You may need to page through the list to see all the parameters.
4. To modify one or more parameters, choose Edit Parameters.
5. In the Edit Parameter Group: screen, scroll using the left and right arrows until you find the binding_protocol parameter, then type ascii in the Value column.
6. Choose Save Changes.
7. To find the name of the parameter you changed, see Memcached Specific Parameters (p. 151).
   If changes to the parameter take place After restart, reboot every cluster that uses this parameter group. For more information, see Rebooting a Cluster (p. 93).

Modifying a Parameter Group (AWS CLI)

To change a parameter's value using the AWS CLI, use the command modify-cache-parameter-group.

Example

To find the name and permitted values of the parameter you want to change, see Memcached Specific Parameters (p. 151)

The following sample code sets the value of two parameters, chunk_size and chunk_size_growth_fact on the parameter group myMem14.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-cache-parameter-group
   --cache-parameter-group-name myMem14
   --parameter-name-values ParameterName=chunk_size,ParameterValue=96
   ParameterName=chunk_size_growth_fact,ParameterValue=1.5
```

For Windows:

```bash
aws elasticache modify-cache-parameter-group
   --cache-parameter-group-name myMem14
   --parameter-name-values ParameterName=chunk_size,ParameterValue=96
   ParameterName=chunk_size_growth_fact,ParameterValue=1.5
```

Output from this command will look something like this.

```json
{
   "CacheParameterGroupName": "myMem14"
}
```

For more information, see modify-cache-parameter-group.
If changes to the parameter take place *After restart*, reboot every cluster that uses this parameter group. For more information, see Rebooting a Cluster (p. 93).

**Modifying a Parameter Group (ElastiCache API)**

To change a parameter group’s parameter values using the ElastiCache API, use the `ModifyCacheParameterGroup` action.

**Example**

To find the name and permitted values of the parameter you want to change, see Memcached Specific Parameters (p. 151)

The following sample code sets the value of two parameters, `chunk_size` and `chunk_size_growth_fact` on the parameter group `myMem14`.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
  ?Action=ModifyCacheParameterGroup
  &CacheParameterGroupName=myMem14
  &ParameterNameValues.member.1.ParameterName=chunk_size
  &ParameterNameValues.member.1.ParameterValue=96
  &ParameterNameValues.member.2.ParameterName=chunk_size_growth_fact
  &ParameterNameValues.member.2.ParameterValue=1.5
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &Version=2015-02-02
  &X-Amz-Credential=<credential>
```

For more information, see `ModifyCacheParameterGroup`.

After updating and saving the parameter, if the change to the parameter take place *After restart*, reboot every cluster that uses this parameter group. For more information, see Rebooting a Cluster (p. 93).
Deleting a Parameter Group

You can delete a custom parameter group using the ElastiCache console, the AWS CLI, or the ElastiCache API.

You cannot delete a parameter group if it is associated with any clusters. Nor can you delete any of the default parameter groups.

Deleting a Parameter Group (Console)

The following procedure shows how to delete a parameter group using the ElastiCache console.

To delete a parameter group using the ElastiCache console

2. To see a list of all available parameter groups, in the left hand navigation pane choose Parameter Groups.
3. Choose the parameter groups you want to delete by choosing the box to the left of the parameter group's name.

   The Delete button will become active.
4. Choose Delete.

   The Delete Parameter Groups confirmation screen will appear.
5. To delete the parameter groups, on the Delete Parameter Groups confirmation screen, choose Delete.

   To keep the parameter groups, choose Cancel.

Deleting a Parameter Group (AWS CLI)

To delete a parameter group using the AWS CLI, use the command delete-cache-parameter-group. For the parameter group to delete, the parameter group specified by --cache-parameter-group-name cannot have any clusters associated with it, nor can it be a default parameter group.

The following sample code deletes the myMem14 parameter group.

Example

For Linux, macOS, or Unix:

```
aws elasticache delete-cache-parameter-group \ 
   --cache-parameter-group-name myMem14
```

For Windows:

```
aws elasticache delete-cache-parameter-group ^
   --cache-parameter-group-name myMem14
```

For more information, see delete-cache-parameter-group.
Deleting a Parameter Group (ElastiCache API)

To delete a parameter group using the ElastiCache API, use the DeleteCacheParameterGroup action. For the parameter group to delete, the parameter group specified by CacheParameterGroupName cannot have any clusters associated with it, nor can it be a default parameter group.

Example

The following sample code deletes the *myMem14* parameter group.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
?Action=DeleteCacheParameterGroup
&CacheParameterGroupName=myMem14
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&Version=2015-02-02
&X-Amz-Credential=<credential>
```

For more information, see [DeleteCacheParameterGroup](#).
Memcached Specific Parameters

If you do not specify a parameter group for your Memcached cluster, then a default parameter group appropriate to your engine version will be used. You can't change the values of any parameters in a default parameter group. However, you can create a custom parameter group and assign it to your cluster at any time. For more information, see Creating a Parameter Group (p. 137).

Topics

- Memcached 1.5.10 Parameter Changes (p. 151)
- Memcached 1.4.34 Added Parameters (p. 152)
- Memcached 1.4.33 Added Parameters (p. 152)
- Memcached 1.4.24 Added Parameters (p. 154)
- Memcached 1.4.14 Added Parameters (p. 155)
- Memcached 1.4.5 Supported Parameters (p. 156)
- Memcached Connection Overhead (p. 158)
- Memcached Node-Type Specific Parameters (p. 158)

Memcached 1.5.10 Parameter Changes

For Memcached 1.5.10, the following additional parameters are supported.

**Parameter group family:** memcached1.5

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>no_modern</td>
<td>Default: 0</td>
<td>An alias for disabling slab_reassign, slab_automove, lru_crawler, lru_maintainer, maxconns_fast commands. No modern also sets the hash_algorithm to jenkins and allows inlining of ASCII VALUE. Applicable to memcached 1.5 and higher. To revert to modern, which is now the default, you must re-launch.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allowed_Values: 0,1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At launch</td>
<td></td>
</tr>
<tr>
<td>inline_ascii_resp</td>
<td>Default: 0</td>
<td>Stores numbers from VALUE response, inside an item, using up to 24 bytes. Small slowdown for ASCII get, faster sets.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allowed_Values: 0,1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At launch</td>
<td></td>
</tr>
</tbody>
</table>

For Memcached 1.5.10, the following parameters are removed.

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>expirezero_does_not_evict</td>
<td>Default: 0</td>
<td>No longer supported in this version.</td>
</tr>
</tbody>
</table>
Memcached Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>modern</td>
<td>Default: 1</td>
<td>Type: boolean</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td>Modifiable: Yes</td>
</tr>
<tr>
<td></td>
<td>(requires re-launch if set to no-modern)</td>
<td>Changes Take Effect: At launch</td>
</tr>
<tr>
<td></td>
<td>Allowed.Values: 0,1</td>
<td>No longer supported in this version. Starting with this version, modern is enabled by default with every launch or re-launch.</td>
</tr>
</tbody>
</table>

Memcached 1.4.34 Added Parameters

For Memcached 1.4.34, no additional parameters are supported.

Parameter group family: memcached1.4

Memcached 1.4.33 Added Parameters

For Memcached 1.4.33, the following additional parameters are supported.

Parameter group family: memcached1.4

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>modern</td>
<td>Default: enabled</td>
<td>An alias to multiple features. Enabling modern is equivalent to turning following commands on and using a murmur3 hash algorithm: slab_reassign, slab_automove, lru_crawler, lru_maintainer, maxconns_fast, and hash_algorithm=murmur3.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td>Changes Take Effect: At launch</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td>watch</td>
<td>Default: enabled</td>
<td>Logs fetches, evictions or mutations. When, for example, user turns watch on, they can see logs when get, set, delete, or update occur.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td>Changes Take Effect: Immediately</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: Immediately</td>
<td>Logs can get dropped if user hits their watcher_logbuf_size and</td>
</tr>
</tbody>
</table>
## Memcached Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>worker_logbuf_size</code></td>
<td>limits.</td>
<td>The minimum number of seconds a client will be allowed to idle before being asked to close. Range of values: 0 to 86400.</td>
</tr>
<tr>
<td><code>idle_timeout</code></td>
<td>Default: 0 (disabled)</td>
<td>The minimum number of seconds a client will be allowed to idle before being asked to close. Range of values: 0 to 86400.</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At Launch</td>
<td></td>
</tr>
<tr>
<td><code>track_sizes</code></td>
<td>Default: disabled</td>
<td>Shows the sizes each slab group has consumed.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At Launch</td>
<td></td>
</tr>
<tr>
<td><code>watcher_logbuf_size</code></td>
<td>Default: 256 (KB)</td>
<td>The watch command turns on stream logging for Memcached. However watch can drop logs if the rate of evictions, mutations or fetches are high enough to cause the logging buffer to become full. In such situations, users can increase the buffer size to reduce the chance of log losses.</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At Launch</td>
<td></td>
</tr>
<tr>
<td><code>worker_logbuf_size</code></td>
<td>Default: 64 (KB)</td>
<td>The watch command turns on stream logging for Memcached. However watch can drop logs if the rate of evictions, mutations or fetches are high enough to cause the logging buffer to become full. In such situations, users can increase the buffer size to reduce the chance of log losses.</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At Launch</td>
<td></td>
</tr>
<tr>
<td><code>slab_chunk_max</code></td>
<td>Default: 524288 (bytes)</td>
<td>Specifies the maximum size of a slab. Setting smaller slab size uses memory more efficiently. Items larger than slab_chunk_max are split over multiple slabs.</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At Launch</td>
<td></td>
</tr>
<tr>
<td><code>lru_crawler metadump</code></td>
<td>Default: disabled</td>
<td>If lru_crawler is enabled this command dumps all keys.</td>
</tr>
<tr>
<td></td>
<td>[all</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: Immediately</td>
<td></td>
</tr>
</tbody>
</table>
# Memcached 1.4.24 Added Parameters

For Memcached 1.4.24, the following additional parameters are supported.

**Parameter group family:** memcached1.4

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>disable_flush_all</td>
<td>Default: 0 (disabled)</td>
<td>Add parameter (<code>-F</code>) to disable flush_all. Useful if you never want to be able to run a full flush on production instances. Values: 0, 1 (user can do a flush_all when the value is 0).</td>
</tr>
<tr>
<td>hash_algorithm</td>
<td>Default: jenkins</td>
<td>The hash algorithm to be used. Permitted values: murmur3 and jenkins.</td>
</tr>
<tr>
<td>lru_crawler</td>
<td>Default: 0 (disabled)</td>
<td>Cleans slab classes of items that have expired. This is a low impact process that runs in the background. Currently requires initiating a crawl using a manual command.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td>Changes Take Effect: After restart</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td>To temporarily enable, run lru_crawler enable at the command line.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lru_crawler 1, 3, 5 crawls slab classes 1, 3, and 5 looking for expired items to add to the freelist. Values: 0, 1</td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: After restart</td>
<td><strong>Note</strong> You can temporarily enable lru_crawler at runtime from the command line. For more information, see the Description column.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note</strong> Enabling lru_crawler at the command line enables the crawler until either disabled at the command line or the next reboot. To enable permanently, you must modify the parameter value. For more information, see Modifying a Parameter Group (p. 146).</td>
</tr>
<tr>
<td>lru_maintainer</td>
<td>Default: 0 (disabled)</td>
<td>A background thread that shuffles items between the LRUs as capacities are reached. Values: 0, 1.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td>Modifiable: Yes</td>
</tr>
<tr>
<td>Name</td>
<td>Details</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: At launch</td>
<td></td>
</tr>
<tr>
<td>expirezero_does_not_evict</td>
<td>Default: 0 (disabled) Type: boolean Modifiable: Yes Changes Take Effect: At launch</td>
<td>When used with lru_maintainer, makes items with an expiration time of 0 unevictable. <strong>Warning</strong> This can crowd out memory available for other evictable items. Can be set to disregard lru_maintainer.</td>
</tr>
</tbody>
</table>

**Memcached 1.4.14 Added Parameters**

For Memcached 1.4.14, the following additional parameters are supported.

**Parameter group family:** memcached1.4

**Parameters added in Memcached 1.4.14**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>config_max</td>
<td>The maximum number of ElastiCache configuration entries.</td>
</tr>
<tr>
<td></td>
<td><strong>Default:</strong> 16 <strong>Type:</strong> integer <strong>Modifiable:</strong> No</td>
</tr>
<tr>
<td>config_size_max</td>
<td>The maximum size of the configuration entries, in bytes.</td>
</tr>
<tr>
<td></td>
<td><strong>Default:</strong> 65536 <strong>Type:</strong> integer <strong>Modifiable:</strong> No</td>
</tr>
<tr>
<td>hashpower_init</td>
<td>The initial size of the ElastiCache hash table, expressed as a power of two.</td>
</tr>
<tr>
<td></td>
<td><strong>Default:</strong> 16 <strong>Type:</strong> integer <strong>Modifiable:</strong> No</td>
</tr>
<tr>
<td>maxconns_fast</td>
<td>The way in which new connections requests are handled when the maximum connection limit is reached. If this parameter is set to 0 (zero), new connections are added to the backlog queue and will wait until other connections are closed. If the parameter is set to 1, ElastiCache sends an error to the client and immediately closes the connection.</td>
</tr>
<tr>
<td></td>
<td><strong>Default:</strong> 0 (false) <strong>Type:</strong> boolean <strong>Modifiable:</strong> Yes</td>
</tr>
</tbody>
</table>
## Memcached Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td>Changes Take Effect: After restart</td>
</tr>
<tr>
<td>slab_automove</td>
<td>Default: 0 (zero)</td>
<td>Adjusts the slab automove algorithm: If this parameter is set to 0 (zero), the automove algorithm is disabled. If it is set to 1, ElastiCache takes a slow, conservative approach to automatically moving slabs. If it is set to 2, ElastiCache aggressively moves slabs whenever there is an eviction. (This mode is not recommended except for testing purposes.) Modifiable: Yes Changes Take Effect: After restart</td>
</tr>
<tr>
<td>slab_reassign</td>
<td>Default: false (false)</td>
<td>Enable or disable slab reassignment. If this parameter is set to 1, you can use the &quot;slabs reassign&quot; command to manually reassign memory. Modifiable: Yes Changes Take Effect: After restart</td>
</tr>
</tbody>
</table>

## Memcached 1.4.5 Supported Parameters

**Parameter group family:** memcached1.4

For Memcached 1.4.5, the following parameters are supported.

### Parameters added in Memcached 1.4.5

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>backlog_queue_limit</td>
<td>Default: 1024</td>
<td>The backlog queue limit.</td>
</tr>
<tr>
<td>binding_protocol</td>
<td>Default: auto</td>
<td>The binding protocol. Permissible values are: ascii and auto. For guidance on modifying the value of binding_protocol, see Modifying a Parameter Group (p. 146).</td>
</tr>
<tr>
<td>cas_disabled</td>
<td>Default: 0 (false)</td>
<td>If 1 (true), check and set (CAS) operations will be disabled, and items stored will consume 8 fewer bytes than with CAS enabled.</td>
</tr>
</tbody>
</table>

API Version 2015-02-02
156
<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>chunk_size</td>
<td>Default: 48</td>
<td>The minimum amount, in bytes, of space to allocate for the smallest item's key, value, and flags.</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: After restart</td>
<td></td>
</tr>
<tr>
<td>chunk_size_growth_factor</td>
<td>Default: 25</td>
<td>The growth factor that controls the size of each successive Memcached chunk; each chunk will be chunk_size_growth_factor times larger than the previous chunk.</td>
</tr>
<tr>
<td></td>
<td>Type: float</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: After restart</td>
<td></td>
</tr>
<tr>
<td>error_on_memory_exhausted</td>
<td>Default: 0 (false)</td>
<td>If 1 (true), when there is no more memory to store items, Memcached will return an error rather than evicting items.</td>
</tr>
<tr>
<td></td>
<td>Type: Boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: After restart</td>
<td></td>
</tr>
<tr>
<td>large_memory_pages</td>
<td>Default: 0 (false)</td>
<td>If 1 (true), ElastiCache will try to use large memory pages.</td>
</tr>
<tr>
<td></td>
<td>Type: Boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td>lock_down_paged_memory</td>
<td>Default: 0 (false)</td>
<td>If 1 (true), ElastiCache will lock down all paged memory.</td>
</tr>
<tr>
<td></td>
<td>Type: Boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td>max_item_size</td>
<td>Default: 1048576</td>
<td>The size, in bytes, of the largest item that can be stored in the cluster.</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: After restart</td>
<td></td>
</tr>
<tr>
<td>max_simultaneous_connections</td>
<td>Default: 600000</td>
<td>The maximum number of simultaneous connections.</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td>maximize_core_file_limit</td>
<td>Default: 0 (false)</td>
<td>If 1 (true), ElastiCache will maximize the core file limit.</td>
</tr>
<tr>
<td></td>
<td>Type: Boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: No</td>
<td></td>
</tr>
</tbody>
</table>
Memcached Connection Overhead

On each node, the memory made available for storing items is the total available memory on that node (which is stored in the `max_cache_memory` parameter) minus the memory used for connections and other overhead (which is stored in the `memcached_connections_overhead` parameter). For example, a node of type `cache.m1.small` has a `max_cache_memory` of 1300MB. With the default `memcached_connections_overhead` value of 100MB, the Memcached process will have 1200MB available to store items.

The default values for the `memcached_connections_overhead` parameter satisfy most use cases; however, the required amount of allocation for connection overhead can vary depending on multiple factors, including request rate, payload size, and the number of connections.

You can change the value of the `memcached_connections_overhead` to better suit the needs of your application. For example, increasing the value of the `memcached_connections_overhead` parameter will reduce the amount of memory available for storing items and provide a larger buffer for connection overhead. Decreasing the value of the `memcached_connections_overhead` parameter will give you more memory to store items, but can increase your risk of swap usage and degraded performance. If you observe swap usage and degraded performance, try increasing the value of the `memcached_connections_overhead` parameter.

**Important**

For the `cache.t1.micro` node type, the value for `memcached_connections_overhead` is determined as follows:

- If you cluster is using the default parameter group, ElastiCache will set the value for `memcached_connections_overhead` to 13MB.
- If your cluster is using a parameter group that you have created yourself, you can set the value of `memcached_connections_overhead` to a value of your choice.

Memcached Node-Type Specific Parameters

Although most parameters have a single value, some parameters have different values depending on the node type used. The following table shows the default values for the `max_cache_memory` and `num_threads` parameters for each node type. The values on these parameters cannot be modified.

<table>
<thead>
<tr>
<th>Node Type</th>
<th>max_cache_memory (in megabytes)</th>
<th>num_threads</th>
</tr>
</thead>
<tbody>
<tr>
<td>cache.t1.micro</td>
<td>213</td>
<td>1</td>
</tr>
<tr>
<td>Node Type</td>
<td>max_cache_memory (in megabytes)</td>
<td>num_threads</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>cache.t2.micro</td>
<td>555</td>
<td>1</td>
</tr>
<tr>
<td>cache.t2.small</td>
<td>1588</td>
<td>1</td>
</tr>
<tr>
<td>cache.t2.medium</td>
<td>3301</td>
<td>2</td>
</tr>
<tr>
<td>cache.t3.micro</td>
<td>512</td>
<td>2</td>
</tr>
<tr>
<td>cache.t3.small</td>
<td>1402</td>
<td>2</td>
</tr>
<tr>
<td>cache.t3.medium</td>
<td>3364</td>
<td>2</td>
</tr>
<tr>
<td>cache.m1.small</td>
<td>3301</td>
<td>1</td>
</tr>
<tr>
<td>cache.m1.medium</td>
<td>3350</td>
<td>1</td>
</tr>
<tr>
<td>cache.m1.large</td>
<td>7100</td>
<td>2</td>
</tr>
<tr>
<td>cache.m1.xlarge</td>
<td>14600</td>
<td>4</td>
</tr>
<tr>
<td>cache.m2.xlarge</td>
<td>33800</td>
<td>2</td>
</tr>
<tr>
<td>cache.m2.2xlarge</td>
<td>30412</td>
<td>4</td>
</tr>
<tr>
<td>cache.m2.4xlarge</td>
<td>68000</td>
<td>16</td>
</tr>
<tr>
<td>cache.m3.medium</td>
<td>2850</td>
<td>1</td>
</tr>
<tr>
<td>cache.m3.large</td>
<td>6200</td>
<td>2</td>
</tr>
<tr>
<td>cache.m3.xlarge</td>
<td>13600</td>
<td>4</td>
</tr>
<tr>
<td>cache.m3.2xlarge</td>
<td>28600</td>
<td>8</td>
</tr>
<tr>
<td>cache.m4.large</td>
<td>6573</td>
<td>2</td>
</tr>
<tr>
<td>cache.m4.xlarge</td>
<td>14618</td>
<td>4</td>
</tr>
<tr>
<td>cache.m4.2xlarge</td>
<td>30412</td>
<td>8</td>
</tr>
<tr>
<td>cache.m4.4xlarge</td>
<td>62234</td>
<td>16</td>
</tr>
<tr>
<td>cache.m4.10xlarge</td>
<td>158355</td>
<td>40</td>
</tr>
<tr>
<td>cache.m5.large</td>
<td>6537</td>
<td>2</td>
</tr>
<tr>
<td>cache.m5.xlarge</td>
<td>13248</td>
<td>4</td>
</tr>
<tr>
<td>cache.m5.2xlarge</td>
<td>26671</td>
<td>8</td>
</tr>
<tr>
<td>cache.m5.4xlarge</td>
<td>53516</td>
<td>16</td>
</tr>
<tr>
<td>cache.m5.12xlarge</td>
<td>160900</td>
<td>48</td>
</tr>
<tr>
<td>cache.m5.24xlarge</td>
<td>321865</td>
<td>96</td>
</tr>
<tr>
<td>cache.c1.xlarge</td>
<td>6600</td>
<td>8</td>
</tr>
<tr>
<td>cache.r3.large</td>
<td>13800</td>
<td>2</td>
</tr>
<tr>
<td>Node Type</td>
<td>max_cache_memory (in megabytes)</td>
<td>num_threads</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>cache.r3.xlarge</td>
<td>29100</td>
<td>4</td>
</tr>
<tr>
<td>cache.r3.2xlarge</td>
<td>59600</td>
<td>8</td>
</tr>
<tr>
<td>cache.r3.4xlarge</td>
<td>120600</td>
<td>16</td>
</tr>
<tr>
<td>cache.r3.8xlarge</td>
<td>120600</td>
<td>32</td>
</tr>
<tr>
<td>cache.r4.large</td>
<td>12590</td>
<td>2</td>
</tr>
<tr>
<td>cache.r4.xlarge</td>
<td>25652</td>
<td>4</td>
</tr>
<tr>
<td>cache.r4.2xlarge</td>
<td>51686</td>
<td>8</td>
</tr>
<tr>
<td>cache.r4.4xlarge</td>
<td>103815</td>
<td>16</td>
</tr>
<tr>
<td>cache.r4.8xlarge</td>
<td>208144</td>
<td>32</td>
</tr>
<tr>
<td>cache.r4.16xlarge</td>
<td>416776</td>
<td>64</td>
</tr>
<tr>
<td>cache.r5.large</td>
<td>13387</td>
<td>2</td>
</tr>
<tr>
<td>cache.r5.xlarge</td>
<td>26953</td>
<td>4</td>
</tr>
<tr>
<td>cache.r5.2xlarge</td>
<td>54084</td>
<td>8</td>
</tr>
<tr>
<td>cache.r5.4xlarge</td>
<td>108347</td>
<td>16</td>
</tr>
<tr>
<td>cache.r5.12xlarge</td>
<td>325400</td>
<td>48</td>
</tr>
<tr>
<td>cache.r5.24xlarge</td>
<td>650869</td>
<td>96</td>
</tr>
</tbody>
</table>

**Note**

All T2 instances are created in an Amazon Virtual Private Cloud (Amazon VPC).
Monitoring Usage, Events, and Costs

To manage your enterprise caching solution, it's important that you know how your clusters are performing and the resources they're consuming. It's also important that you know the events that are being generated and the costs of your deployment.

Amazon CloudWatch provides metrics for monitoring your cache performance. In addition, cost allocation tags help you monitor and manage costs.

Topics
- Monitoring Use with Metrics (p. 162)
- Monitoring ElastiCache Events (p. 170)
- Monitoring Costs with Cost Allocation Tags (p. 179)
- Managing Costs with Reserved Nodes (p. 186)
Monitoring Use with Metrics

provides metrics that enable you to monitor your clusters. You can access these metrics through . For more information on , see the documentation.

provides both host-level metrics (for example, CPU usage) and metrics that are specific to the cache engine software (for example, cache gets and cache misses). These metrics are measured and published for each Cache node in 60-second intervals.

Important
You should consider setting alarms on certain key metrics, so that you will be notified if your cache cluster's performance starts to degrade. For more information, see Which Metrics Should I Monitor? (p. 166) in this guide.

Topics
• Host-Level Metrics (p. 162)
• Metrics for Memcached (p. 163)
• Which Metrics Should I Monitor? (p. 166)
• Choosing Metric Statistics and Periods (p. 167)
• Monitoring CloudWatch Cluster and Node Metrics (p. 167)

Host-Level Metrics

The AWS/ElastiCache namespace includes the following host-level metrics for individual cache nodes.

See Also
• Metrics for Memcached (p. 163)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUUtilization</td>
<td>The percentage of CPU utilization.</td>
<td>Percent</td>
</tr>
<tr>
<td>FreeableMemory</td>
<td>The amount of free memory available on the host.</td>
<td>Bytes</td>
</tr>
<tr>
<td>NetworkBytesIn</td>
<td>The number of bytes the host has read from the network.</td>
<td>Bytes</td>
</tr>
<tr>
<td>NetworkBytesOut</td>
<td>The number of bytes the host has written to the network.</td>
<td>Bytes</td>
</tr>
<tr>
<td>NetworkPacketsIn</td>
<td>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.</td>
<td>Count</td>
</tr>
<tr>
<td>NetworkPacketsOut</td>
<td>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.</td>
<td>Count</td>
</tr>
<tr>
<td>SwapUsage</td>
<td>The amount of swap used on the host.</td>
<td>Count</td>
</tr>
</tbody>
</table>
Metrics for Memcached

The AWS/ElastiCache namespace includes the following Memcached metrics.

The AWS/ElastiCache namespace includes the following metrics that are derived from the Memcached stats command. Each metric is calculated at the cache node level.

For complete documentation of the Memcached info command, see .

See Also

- Host-Level Metrics (p. 162)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>BytesReadIntoMemcached</td>
<td>The number of bytes that have been read from the network by the cache node.</td>
<td>Bytes</td>
</tr>
<tr>
<td>BytesUsedForCache</td>
<td>The number of bytes used to store cache items.</td>
<td>Bytes</td>
</tr>
<tr>
<td>BytesWrittenOutFromMemcached</td>
<td>The number of bytes that have been written to the network by the cache node.</td>
<td>Bytes</td>
</tr>
<tr>
<td>CasBadval</td>
<td>The number of CAS (check and set) requests the cache has received where the Cas value did not match the Cas value stored.</td>
<td>Count</td>
</tr>
<tr>
<td>CasHits</td>
<td>The number of Cas requests the cache has received where the requested key was found and the Cas value matched.</td>
<td>Count</td>
</tr>
<tr>
<td>CasMisses</td>
<td>The number of Cas requests the cache has received where the key requested was not found.</td>
<td>Count</td>
</tr>
<tr>
<td>CmdFlush</td>
<td>The number of flush commands the cache has received.</td>
<td>Count</td>
</tr>
<tr>
<td>CmdGets</td>
<td>The number of get commands the cache has received.</td>
<td>Count</td>
</tr>
<tr>
<td>CmdSet</td>
<td>The number of set commands the cache has received.</td>
<td>Count</td>
</tr>
<tr>
<td>CurrConnections</td>
<td>A count of the number of connections connected to the cache at an instant in time. ElastiCache uses two to three of the connections to monitor the cluster. In addition to the above, memcached creates a number of internal connections equal to twice the threads used for the node type. The thread count for the various node types can be see in the Nodetype Specific Parameters of the applicable Parameter Group. The total connections is the sum of client connections, the connections for monitoring and the internal connections mentioned above.</td>
<td>Count</td>
</tr>
</tbody>
</table>

API Version 2015-02-02
# Metrics for Memcached

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>CurrItems</td>
<td>A count of the number of items currently stored in the cache.</td>
<td>Count</td>
</tr>
<tr>
<td>DecrHits</td>
<td>The number of decrement requests the cache has received where the requested key was found.</td>
<td>Count</td>
</tr>
<tr>
<td>DecrMisses</td>
<td>The number of decrement requests the cache has received where the requested key was not found.</td>
<td>Count</td>
</tr>
<tr>
<td>DeleteHits</td>
<td>The number of delete requests the cache has received where the requested key was found.</td>
<td>Count</td>
</tr>
<tr>
<td>DeleteMisses</td>
<td>The number of delete requests the cache has received where the requested key was not found.</td>
<td>Count</td>
</tr>
<tr>
<td>Evictions</td>
<td>The number of non-expired items the cache evicted to allow space for new writes.</td>
<td>Count</td>
</tr>
<tr>
<td>GetHits</td>
<td>The number of get requests the cache has received where the key requested was found.</td>
<td>Count</td>
</tr>
<tr>
<td>GetMisses</td>
<td>The number of get requests the cache has received where the key requested was not found.</td>
<td>Count</td>
</tr>
<tr>
<td>IncrHits</td>
<td>The number of increment requests the cache has received where the key requested was found.</td>
<td>Count</td>
</tr>
<tr>
<td>IncrMisses</td>
<td>The number of increment requests the cache has received where the key requested was not found.</td>
<td>Count</td>
</tr>
<tr>
<td>Reclaimed</td>
<td>The number of expired items the cache evicted to allow space for new writes.</td>
<td>Count</td>
</tr>
</tbody>
</table>

For Memcached 1.4.14, the following additional metrics are provided.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>BytesUsedForHash</td>
<td>The number of bytes currently used by hash tables.</td>
<td>Bytes</td>
</tr>
<tr>
<td>CmdConfigGet</td>
<td>The cumulative number of config get requests.</td>
<td>Count</td>
</tr>
<tr>
<td>CmdConfigSet</td>
<td>The cumulative number of config set requests.</td>
<td>Count</td>
</tr>
<tr>
<td>CmdTouch</td>
<td>The cumulative number of touch requests.</td>
<td>Count</td>
</tr>
<tr>
<td>CurrConfig</td>
<td>The current number of configurations stored.</td>
<td>Count</td>
</tr>
<tr>
<td>EvictedUnfetched</td>
<td>The number of valid items evicted from the least recently used cache (LRU) which were never touched after being set.</td>
<td>Count</td>
</tr>
<tr>
<td>ExpiredUnfetched</td>
<td>The number of expired items reclaimed from the LRU which were never touched after being set.</td>
<td>Count</td>
</tr>
<tr>
<td>SlabsMoved</td>
<td>The total number of slab pages that have been moved.</td>
<td>Count</td>
</tr>
</tbody>
</table>

API Version 2015-02-02

164
## Metrics for Memcached

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>TouchHits</td>
<td>The number of keys that have been touched and were given a new expiration time.</td>
<td>Count</td>
</tr>
<tr>
<td>TouchMisses</td>
<td>The number of items that have been touched, but were not found.</td>
<td>Count</td>
</tr>
</tbody>
</table>

The AWS/ElastiCache namespace includes the following calculated cache-level metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>NewConnections</td>
<td>The number of new connections the cache has received. This is derived from the memcached total_connections statistic by recording the change in total_connections across a period of time. This will always be at least 1, due to a connection reserved for ElastiCache.</td>
<td>Count</td>
</tr>
<tr>
<td>NewItems</td>
<td>The number of new items the cache has stored. This is derived from the memcached total_items statistic by recording the change in total_items across a period of time.</td>
<td>Count</td>
</tr>
<tr>
<td>UnusedMemory</td>
<td>The amount of memory not used by data. This is derived from the Memcached statistics limit_maxbytes and bytes by subtracting bytes from limit_maxbytes. Because Memcached overhead uses memory in addition to that used by data, UnusedMemory should not be considered to be the amount of memory available for additional data. You may experience evictions even though you still have some unused memory. For more detailed information, see Memcached item memory usage.</td>
<td>Bytes</td>
</tr>
</tbody>
</table>
Which Metrics Should I Monitor?

The following metrics offer good insight into performance. In most cases, we recommend that you set alarms for these metrics so that you can take corrective action before performance issues occur.

**Metrics to Monitor**

- CPUUtilization (p. 166)
- SwapUsage (p. 166)
- Evictions (p. 166)
- CurrConnections (p. 166)

**CPUUtilization**

This is a host-level metric reported as a percent. For more information, see Host-Level Metrics (p. 162).

Because Memcached is multi-threaded, this metric can be as high as 90%. If you exceed this threshold, scale your cache cluster up by using a larger cache node type, or scale out by adding more cache nodes.

**SwapUsage**

This is a host-level metric reported in bytes. For more information, see Host-Level Metrics (p. 162).

This metric should not exceed 50 MB. If it does, we recommend that you increase the ConnectionOverhead parameter value.

**Evictions**

This is a cache engine metric. We recommend that you determine your own alarm threshold for this metric based on your application needs.

If you exceed your chosen threshold, scale your cluster up by using a larger node type, or scale out by adding more nodes.

**CurrConnections**

This is a cache engine metric. We recommend that you determine your own alarm threshold for this metric based on your application needs.

An increasing number of CurrConnections might indicate a problem with your application; you will need to investigate the application behavior to address this issue.
Choosing Metric Statistics and Periods

While will allow you to choose any statistic and period for each metric, not all combinations will be useful. For example, the Average, Minimum, and Maximum statistics for CPUUtilization are useful, but the Sum statistic is not.

All samples are published for a 60 second duration for each individual cache node. For any 60 second period, a cache node metric will only contain a single sample.

For further information on how to retrieve metrics for your cache nodes, see Monitoring CloudWatch Cluster and Node Metrics (p. 167).

Monitoring CloudWatch Cluster and Node Metrics

ElastiCache and CloudWatch are integrated so you can gather a variety of metrics. You can monitor these metrics using CloudWatch.

Note
The following examples require the CloudWatch command line tools. For more information about CloudWatch and to download the developer tools, see the CloudWatch product page.

The following procedures show you how to use CloudWatch to gather storage space statistics for an cache cluster for the past hour.

Note
The StartTime and EndTime values supplied in the examples below are for illustrative purposes. You must substitute appropriate start and end time values for your cache nodes.

For information on ElastiCache limits, see AWS Service Limits for ElastiCache.

Monitoring CloudWatch Cluster and Node Metrics (Console)

To gather CPU utilization statistics for a cache cluster

2. Select the cache nodes you want to view metrics for.

   Note
   Selecting more than 20 nodes disables viewing metrics on the console.
   a. On the Cache Clusters page of the AWS Management Console, click the name of one or more cache clusters.

   The detail page for the cache cluster appears.
   b. Click the Nodes tab at the top of the window.
   c. On the Nodes tab of the detail window, select the cache nodes that you want to view metrics for.

   A list of available CloudWatch Metrics appears at the bottom of the console window.
   d. Click on the CPU Utilization metric.

   The CloudWatch console will open, displaying your selected metrics. You can use the Statistic and Period drop-down list boxes and Time Range tab to change the metrics being displayed.
Monitoring CloudWatch Cluster and Node Metrics Using the CloudWatch CLI

To gather CPU utilization statistics for a cache cluster

- For Linux, macOS, or Unix:

  ```bash
  aws cloudwatch get-metric-statistics \
  --namespace AWS/ElastiCache \
  --metric-name CPUUtilization \
  --dimensions="Name=CacheClusterId=mycachecluster,Value=CacheNodeId=0002" \
  --statistics=Average \
  --start-time 2018-07-05T00:00:00 \
  --end-time 2018-07-06T00:00:00 \
  --period=3600
  ```

  For Windows:

  ```bash
  aws cloudwatch get-metric-statistics ^
  --namespace AWS/ElastiCache ^
  --metric-name CPUUtilization ^
  --dimensions="Name=CacheClusterId=mycachecluster,Value=CacheNodeId=0002" ^
  --statistics=Average ^
  --start-time 2018-07-05T00:00:00 ^
  --end-time 2018-07-06T00:00:00 ^
  --period=3600
  ```

Monitoring CloudWatch Cluster and Node Metrics Using the CloudWatch API

To gather CPU utilization statistics for a cache cluster

- Call the CloudWatch API `GetMetricStatistics` with the following parameters (note that the start and end times are shown as examples only; you will need to substitute your own appropriate start and end times):

  - Statistics.member.1=Average
  - Namespace=AWS/ElastiCache
  - StartTime=2013-07-05T00:00:00
  - EndTime=2013-07-06T00:00:00
  - Period=60
  - MeasureName=CPUUtilization
  - Dimensions=CacheClusterId=mycachecluster,CacheNodeid=0002

  **Example**

  ```text
  http://monitoring.amazonaws.com/\n  ?Action=GetMetricStatistics\n  &SignatureVersion=4\n  &Version=2014-12-01\n  &StartTime=2018-07-05T00:00:00\n  &EndTime=2018-07-06T23:59:00\n  &Period=3600
  ```

API Version 2015-02-02
<table>
<thead>
<tr>
<th>&amp;Statistics.member.1=Average</th>
<th>&amp;Dimensions.member.1=&quot;CacheClusterId=mycachecluster&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;Dimensions.member.2=&quot;CacheNodeId=0002&quot;</td>
<td>&amp;Namespace=AWS/ElastiCache</td>
</tr>
<tr>
<td>&amp;MeasureName=CPUUtilization</td>
<td>&amp;Timestamp=2018-07-07T17:48:21.746Z</td>
</tr>
<tr>
<td>&amp;AWSAccessKeyId=&lt;AWS Access Key ID&gt;</td>
<td>&amp;Signature=&lt;Signature&gt;</td>
</tr>
</tbody>
</table>
Monitoring ElastiCache Events

When significant events happen for a cluster, ElastiCache sends notification to a specific Amazon SNS topic. Examples include a failure to add a node, success in adding a node, the modification of a security group, and others. By monitoring for key events, you can know the current state of your clusters and, depending upon the event, be able to take corrective action.

Topics
- Managing ElastiCache Amazon SNS Notifications (p. 170)
- Viewing ElastiCache Events (p. 173)
- Event Notifications and Amazon SNS (p. 175)

Managing ElastiCache Amazon SNS Notifications

You can configure ElastiCache to send notifications for important cluster events using Amazon Simple Notification Service (Amazon SNS). In these examples, you will configure a cluster with the Amazon Resource Name (ARN) of an Amazon SNS topic to receive notifications.

Note
This topic assumes that you've signed up for Amazon SNS and have set up and subscribed to an Amazon SNS topic. For information on how to do this, see the Amazon Simple Notification Service Developer Guide.

Adding an Amazon SNS Topic

The following sections show you how to add an Amazon SNS topic using the AWS Console, the AWS CLI, or the ElastiCache API.

Adding an Amazon SNS Topic (Console)

The following procedure shows you how to add an Amazon SNS topic for a cluster.

Note
This process can also be used to modify the Amazon SNS topic.

To add or modify an Amazon SNS topic for a cluster (Console)

2. In Clusters, choose the cluster for which you want to add or modify an Amazon SNS topic ARN.
3. Choose Modify.
4. In Modify Cluster under Topic for SNS Notification, choose the SNS topic you want to add, or choose Manual ARN input and type the ARN of the Amazon SNS topic.
5. Choose Modify.

Adding an Amazon SNS Topic (AWS CLI)

To add or modify an Amazon SNS topic for a cluster, use the AWS CLI command modify-cache-cluster.

The following code example adds an Amazon SNS topic arn to my-cluster.

For Linux, macOS, or Unix:
Managing ElastiCache Amazon SNS Notifications

For more information, see modify-cache-cluster.

Adding an Amazon SNS Topic (ElastiCache API)

To add or modify an Amazon SNS topic for a cluster, call the ModifyCacheCluster action with the following parameters:

- **CacheClusterId=my-cluster**
- **TopicArn=arn%3Aaws%3Asns%3Aus-west-2%3A565419523791%3AElastiCacheNotifications**

Example

```
https://elasticache.amazon.com/
?Action=ModifyCacheCluster
&ApplyImmediately=false
&CacheClusterId=my-cluster
&NotificationTopicArn=arn%3Aaws%3Asns%3Aus-west-2%3A565419523791%3AElastiCacheNotifications
&Version=2014-12-01
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&X-Amz-Algorithm=AWS4-HMAC-SHA256
&X-Amz-Date=20141201T220302Z
&X-Amz-SignedHeaders=Host
&X-Amz-Expires=20141201T220302Z
&X-Amz-Credential=<credential>
&X-Amz-Signature=<signature>
```

For more information, see ModifyCacheCluster.

Enabling and Disabling Amazon SNS Notifications

You can turn notifications on or off for a cluster. The following procedures show you how to disable Amazon SNS notifications.

Enabling and Disabling Amazon SNS Notifications (Console)

To disable Amazon SNS notifications using the AWS Management Console

2. To see a list of your clusters running Memcached, in the navigation pane choose **Memcached**.
3. Choose the box to the left of the cluster you want to modify notification for.
4. Choose **Modify**.
5. In **Modify Cluster** under **Topic for SNS Notification**, choose **Disable Notifications**.
6. Choose **Modify**.

### Enabling and Disabling Amazon SNS Notifications (AWS CLI)

To disable Amazon SNS notifications, use the command `modify-cache-cluster` with the following parameters:

For Linux, macOS, or Unix:

```bash
aws elasticache modify-cache-cluster \
  --cache-cluster-id my-cluster \
  --notification-topic-status inactive
```

For Windows:

```bash
aws elasticache modify-cache-cluster ^
  --cache-cluster-id my-cluster ^
  --notification-topic-status inactive
```

### Enabling and Disabling Amazon SNS Notifications (ElastiCache API)

To disable Amazon SNS notifications, call the `ModifyCacheCluster` action with the following parameters:

- CacheClusterId=my-cluster
- NotificationTopicStatus=inactive

This call returns output similar to the following:

**Example**

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=ModifyCacheCluster
  &ApplyImmediately=false
  &CacheClusterId=my-cluster
  &NotificationTopicStatus=inactive
  &Version=2014-12-01
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20141201T220302Z
  &X-Amz-Algorithm=AWS4-HMAC-SHA256
  &X-Amz-Date=20141201T220302Z
  &X-Amz-SignedHeaders=Host
  &X-Amz-Expires=20141201T220302Z
  &X-Amz-Credential=<credential>
  &X-Amz-Signature=<signature>
```
Viewing ElastiCache Events

ElastiCache logs events that relate to your cluster instances, security groups, and parameter groups. This information includes the date and time of the event, the source name and source type of the event, and a description of the event. You can easily retrieve events from the log using the ElastiCache console, the AWS CLI `describe-events` command, or the ElastiCache API action `DescribeEvents`.

The following procedures show you how to view all ElastiCache events for the past 24 hours (1440 minutes).

Viewing ElastiCache Events (Console)

The following procedure displays events using the ElastiCache console.

To view events using the ElastiCache console

2. To see a list of all available events, in the navigation pane, choose Events.

On the Events screen each row of the list represents one event and displays the event source, the event type (cache-cluster, cache-parameter-group, cache-security-group, or cache-subnet-group), the GMT time of the event, and a description of the event.

Using the Filter you can specify whether you want to see all events, or just events of a specific type in the event list.

Viewing ElastiCache Events (AWS CLI)

To generate a list of ElastiCache events using the AWS CLI, use the command `describe-events`. You can use optional parameters to control the type of events listed, the time frame of the events listed, the maximum number of events to list, and more.

The following code lists up to 40 cache cluster events.

```bash
aws elasticache describe-events --source-type cache-cluster --max-items 40
```

The following code lists all events for the past 24 hours (1440 minutes).

```bash
aws elasticache describe-events --duration 1440
```

The output from the `describe-events` command looks something like this.

```json
{
    "Events": [
    {
        "Date": "2017-03-29T22:17:37.781Z",
        "Message": "Added cache node 0001 in Availability Zone us-west-2a",
        "SourceIdentifier": "mem01",
        "SourceType": "cache-cluster"
    },
    {
        "Date": "2017-03-29T22:17:37.769Z",
        "Message": "Cache cluster created",
        "SourceIdentifier": "mem01",
        "SourceType": "cache-cluster"
    }
    ]
}
```
For more information, such as available parameters and permitted parameter values, see `describe-events`.

**Viewing ElastiCache Events (ElastiCache API)**

To generate a list of ElastiCache events using the ElastiCache API, use the `DescribeEvents` action. You can use optional parameters to control the type of events listed, the time frame of the events listed, the maximum number of events to list, and more.

The following code lists the 40 most recent cache-cluster events.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeEvents
&MaxRecords=40
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&SourceType=cache-cluster
&Timestamp=20150202T192317Z
&Version=2015-02-02
&X-Amz-Credential=<credential>
```

The following code lists the cache-cluster events for the past 24 hours (1440 minutes).

```plaintext
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeEvents
&Duration=1440
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&SourceType=cache-cluster
&Timestamp=20150202T192317Z
&Version=2015-02-02
&X-Amz-Credential=<credential>
```

The above actions should produce output similar to the following.

```xml
<DescribeEventsResponse xmlns="http://elasticache.amazonaws.com/doc/2015-02-02/">
  <DescribeEventsResult>
    <Events>
      <Event>
        <Message>Cache cluster created</Message>
        <SourceType>cache-cluster</SourceType>
        <Date>2015-02-02T18:22:18.202Z</Date>
        <SourceIdentifier>mem01</SourceIdentifier>
      </Event>
    </Events>
  </DescribeEventsResult>
  <ResponseMetadata>
    <RequestId>e21c81b4-b9cd-11e3-8a16-7978bb24ffdf</RequestId>
  </ResponseMetadata>
</DescribeEventsResponse>
```

For more information, such as available parameters and permitted parameter values, see `DescribeEvents`.
Event Notifications and Amazon SNS

ElastiCache can publish messages using Amazon Simple Notification Service (SNS) when significant events happen on a cache cluster. This feature can be used to refresh the server-lists on client machines connected to individual cache node endpoints of a cache cluster.

Note
For more information on Amazon Simple Notification Service (SNS), including information on pricing and links to the Amazon SNS documentation, see the Amazon SNS product page.

Notifications are published to a specified Amazon SNS topic. The following are requirements for notifications:

- Only one topic can be configured for ElastiCache notifications.
- The AWS account that owns the Amazon SNS topic must be the same account that owns the cache cluster on which notifications are enabled.

Example ElastiCache SNS Notification

The following example shows an ElastiCache Amazon SNS notification for successfully creating a cache cluster.

Example

```
{
    "Date": "2015-12-05T01:02:18.336Z",
    "Message": "Cache cluster created",
    "SourceIdentifier": "memcache-ni",
    "SourceType": "cache-cluster"
}
```

ElastiCache Events

The following ElastiCache events trigger Amazon SNS notifications:

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Message</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElastiCache:AddCacheNodeComplete</td>
<td>cluster-name</td>
<td>A cache node has been added to the cache cluster and is ready for use.</td>
</tr>
<tr>
<td>ElastiCache:AddCacheNodeFailed due to insufficient free IP addresses</td>
<td>cluster-name</td>
<td>A cache node could not be added because there are not enough available IP addresses.</td>
</tr>
<tr>
<td>ElastiCache:CacheClusterParametersChanged</td>
<td>cluster-name</td>
<td>One or more cache cluster parameters have been changed.</td>
</tr>
<tr>
<td>ElastiCache:CacheClusterProvisioningComplete</td>
<td>cluster-name</td>
<td>The provisioning of a cache cluster is completed, and the cache nodes in the cache cluster are ready to use.</td>
</tr>
<tr>
<td>ElastiCache:CacheClusterProvisioningFailed due to incompatible network state</td>
<td>cluster-name</td>
<td>An attempt was made to launch a new cache cluster into a nonexistent virtual private cloud (VPC).</td>
</tr>
<tr>
<td>Event Name</td>
<td>Message</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ElastiCache:CacheClusterScalingComplete</td>
<td>cluster-name</td>
<td>Scale up for cache-cluster completed successfully.</td>
</tr>
<tr>
<td>ElastiCache:CacheClusterScalingFailed</td>
<td>cluster-name</td>
<td>Scale-up operation on cache-cluster failed.</td>
</tr>
<tr>
<td>ElastiCache:CacheClusterSecurityGroupModified</td>
<td>cluster-name</td>
<td>One of the following events has occurred:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The list of cache security groups authorized for the cache cluster has been modified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• One or more new EC2 security groups have been authorized on any of the cache security groups associated with the cache cluster.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• One or more EC2 security groups have been revoked from any of the cache security groups associated with the cache cluster.</td>
</tr>
<tr>
<td>ElastiCache:CacheNodeReplaceStarted</td>
<td>cluster-name</td>
<td>ElastiCache has detected that the host running a cache node is degraded or unreachable and has started replacing the cache node.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The DNS entry for the replaced cache node is not changed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In most instances, you do not need to refresh the server-list for your clients when this event occurs. However, some cache client libraries may stop using the cache node even after ElastiCache has replaced the cache node; in this case, the application should refresh the server-list when this event occurs.</td>
</tr>
</tbody>
</table>
### Event Notifications and Amazon SNS

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Message</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElastiCache:CacheNodeReplaceComplete &lt;cluster-name&gt;</td>
<td></td>
<td>ElastiCache has detected that the host running a cache node is degraded or unreachable and has completed replacing the cache node.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>The DNS entry for the replaced cache node is not changed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In most instances, you do not need to refresh the server-list for your clients when this event occurs. However, some cache client libraries may stop using the cache node even after ElastiCache has replaced the cache node; in this case, the application should refresh the server-list when this event occurs.</td>
</tr>
<tr>
<td>ElastiCache:CacheNodesRebooted &lt;cluster-name&gt;</td>
<td></td>
<td>One or more cache nodes has been rebooted.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Message (Memcached): &quot;Cache node %s shutdown&quot; Then a second message: &quot;Cache node %s restarted&quot;</td>
</tr>
<tr>
<td>ElastiCache:DeleteCacheClusterComplete &lt;cluster-name&gt;</td>
<td></td>
<td>The deletion of a cache cluster and all associated cache nodes has completed.</td>
</tr>
<tr>
<td>ElastiCache:FailoverComplete &lt;cluster-name&gt;</td>
<td></td>
<td>Failover over to a replica node was successful.</td>
</tr>
<tr>
<td>(Redis only)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ElastiCache:NodeReplacementCanceled &lt;cluster-name&gt;</td>
<td></td>
<td>A node in your cluster that was scheduled for replacement is no longer scheduled for replacement.</td>
</tr>
<tr>
<td>ElastiCache:NodeReplacementRescheduled &lt;cluster-name&gt;</td>
<td></td>
<td>A node in your cluster previously scheduled for replacement has been rescheduled for replacement during the new window described in the notification.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>For information on what actions you can take, see Replacing Nodes (p. 50).</td>
</tr>
</tbody>
</table>
### Event Notifications and Amazon SNS

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Message</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElastiCache:NodeReplacementScheduled</td>
<td>cluser-name</td>
<td>A node in your cluster is scheduled for replacement during the window described in the notification. For information on what actions you can take, see Replacing Nodes (p. 50).</td>
</tr>
<tr>
<td>ElastiCache:RemoveCacheNodeComplete</td>
<td>cluser-name</td>
<td>A cache node has been removed from the cache cluster.</td>
</tr>
<tr>
<td>ElastiCache:SnapshotComplete (Redis only)</td>
<td>cluser-name</td>
<td>A cache snapshot has completed successfully.</td>
</tr>
<tr>
<td>ElastiCache:SnapshotFailed (Redis only)</td>
<td>cluser-name</td>
<td>A cache snapshot has failed. See the cluster’s cache events for more a detailed cause. If you describe the snapshot, see DescribeSnapshots, the status will be failed.</td>
</tr>
</tbody>
</table>

**Related topics**

- Viewing ElastiCache Events (p. 173)
Monitoring Costs with Cost Allocation Tags

When you add cost allocation tags to your resources in Amazon ElastiCache, you can track costs by grouping expenses on your invoices by resource tag values.

An ElastiCache cost allocation tag is a key-value pair that you define and associate with an ElastiCache resource. The key and value are case-sensitive. You can use a tag key to define a category, and the tag value can be an item in that category. For example, you might define a tag key of CostCenter and a tag value of 10010, indicating that the resource is assigned to the 10010 cost center. You can also use tags to designate resources as being used for test or production by using a key such as Environment and values such as test or production. We recommend that you use a consistent set of tag keys to make it easier to track costs associated with your resources.

Use cost allocation 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. Then, to see the cost of combined resources, organize your billing information according to resources with 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.

You can also combine tags to track costs at a greater level of detail. For example, to track your service costs by region you might use the tag keys Service and Region. On one resource you might have the values ElastiCache and Asia Pacific (Singapore), and on another resource the values ElastiCache and Europe (Frankfurt). You can then see your total ElastiCache costs broken out by region. For more information, see Use Cost Allocation Tags in the AWS Billing and Cost Management User Guide.

You can add ElastiCache cost allocation tags to Memcached clusters. When you add, list, modify, copy, or remove a tag, the operation is applied only to the specified cluster.

Characteristics of ElastiCache cost allocation tags

- Cost allocation tags are applied to ElastiCache resources which are specified in CLI and API operations as an ARN. The resource-type will be a "cluster".

  Sample ARN: arn:aws:elasticache:<region>::<customer-id>::<resource-type>::<resource-name>

  Memcached: Tags are only applied to clusters.


- The tag key is the required name of the tag. The key's string value can be from 1 to 128 Unicode characters long and cannot be prefixed with aws:. The string can contain only the set of Unicode letters, digits, blank spaces, underscores (_), periods (.), colons (:), backslashes (\), equal signs (=), plus signs (+), hyphens (-), or at signs (@).

- The tag value is the optional value of the tag. The value's string value can be from 1 to 256 Unicode characters in length and cannot be prefixed with aws:. The string can contain only the set of Unicode letters, digits, blank spaces, underscores (_), periods (.), colons (:), backslashes (\), equal signs (=), plus signs (+), hyphens (-), or at signs (@).

- An ElastiCache resource can have a maximum of 50 tags.

- Values do not have to be unique in a tag set. For example, you can have a tag set where the keys Service and Application both have the value ElastiCache.
AWS does not apply any semantic meaning to your tags. Tags are interpreted strictly as character strings. AWS does not automatically set any tags on any ElastiCache resource.

You can add, list, modify, or remove tags from an ElastiCache resource by using the ElastiCache management console, AWS CLI, or ElastiCache API.

Topics

- Managing Your Tags Using the ElastiCache Console (p. 180)
- Managing Your Cost Allocation Tags Using the AWS CLI (p. 181)
- Managing Your Cost Allocation Tags Using the ElastiCache API (p. 183)

Managing Your Tags Using the ElastiCache Console

You can use the Amazon ElastiCache console to add, modify, or remove cost allocation tags.

The following procedure walks you through viewing, adding, modifying, or deleting one or more cost allocation tags using the ElastiCache management console.

Managing tags on a Memcached cluster using the AWS Management Console

2. Choose Memcached.
3. Choose the box to the left of the cluster's name you want to add tags to.
4. Choose Manage Tags, and then use the dialog box to manage your tags.

5. For each tag you want to add, modify, or remove:

   **To add, modify, or remove tags**

   - **To add a tag:** In the Key column, type a key name in the box that displays Add key and an optional value in the box to the right of the key name.
   - **To modify a tag:** In the Value column, type a new value or remove the existing value for the tag.
   - **To remove a tag:** Choose the X to the right of the tag.

6. When you're finished, choose Apply Changes.
Managing Your Cost Allocation Tags Using the AWS CLI

You can use the AWS CLI to add, modify, or remove cost allocation tags.

Cost allocation tags are applied to ElastiCache for Memcached clusters. The cluster to be tagged is specified using an ARN (Amazon Resource Name).


Topics
- Listing Tags Using the AWS CLI (p. 181)
- Adding Tags Using the AWS CLI (p. 182)
- Modifying Tags Using the AWS CLI (p. 182)
- Removing Tags Using the AWS CLI (p. 183)

Listing Tags Using the AWS CLI

You can use the AWS CLI to list tags on an existing ElastiCache resource by using the list-tags-for-resource operation.

The following code uses the AWS CLI to list the tags on the Memcached cluster my-cluster in the us-west-2 region.

For Linux, macOS, or Unix:

```bash
```

For Windows:

```bash
```

Output from this operation will look something like the following, a list of all the tags on the resource.

```json
{
  "TagList": [
    {
      "Value": "10110",
      "Key": "CostCenter"
    },
    {
      "Value": "EC2",
      "Key": "Service"
    }
  ]
}
```

If there are no tags on the resource, the output will be an empty TagList.

```json
{
  "TagList": []
}
```
Adding Tags Using the AWS CLI

You can use the AWS CLI to add tags to an existing ElastiCache resource by using the `add-tags-to-resource` CLI operation. If the tag key does not exist on the resource, the key and value are added to the resource. If the key already exists on the resource, the value associated with that key is updated to the new value.

The following code uses the AWS CLI to add the keys `Service` and `Region` with the values `elasticache` and `us-west-2` respectively to the cluster `my-cluster` in region `us-west-2`.

For Linux, macOS, or Unix:

```bash
aws elasticache add-tags-to-resource
  --tags Key=Service,Value=elasticache
  Key=Region,Value=us-west-2
```

For Windows:

```bash
aws elasticache add-tags-to-resource
  --tags Key=Service,Value=elasticache
  Key=Region,Value=us-west-2
```

Output from this operation will look something like the following, a list of all the tags on the resource following the operation.

```json
{
  "TagList": [
    {
      "Value": "elasticache",
      "Key": "Service"
    },
    {
      "Value": "us-west-2",
      "Key": "Region"
    }
  ]
}
```

For more information, see the AWS CLI for ElastiCache `list-tags-for-resource`.

You can also use the AWS CLI to add tags to a cluster when you create a new cluster by using the operation `create-cache-cluster`. You cannot add tags when creating a cluster using the ElastiCache management console. After the cluster is created, you can then use the console to add tags to the cluster.

Modifying Tags Using the AWS CLI

You can use the AWS CLI to modify the tags on an ElastiCache Memcached cluster.

To modify tags:

- Use `add-tags-to-resource` to either add a new tag and value or to change the value associated with an existing tag.
• Use `remove-tags-from-resource` to remove specified tags from the resource.

Output from either operation will be a list of tags and their values on the specified cluster.

**Removing Tags Using the AWS CLI**

You can use the AWS CLI to remove tags from an existing ElastiCache for Memcached cluster by using the `remove-tags-from-resource` operation.

The following code uses the AWS CLI to remove the tags with the keys `Service` and `Region` from the cluster `my-cluster` in the `us-west-2` region.

For Linux, macOS, or Unix:

```
aws elasticache remove-tags-from-resource \
  --tag-keys PM Service
```

For Windows:

```
aws elasticache remove-tags-from-resource ^
  --tag-keys PM Service
```

Output from this operation will look something like the following, a list of all the tags on the resource following the operation.

```json
{
  "TagList": []
}
```

For more information, see the AWS CLI for ElastiCache `remove-tags-from-resource`.

**Managing Your Cost Allocation Tags Using the ElastiCache API**

You can use the ElastiCache API to add, modify, or remove cost allocation tags.

Cost allocation tags are applied to ElastiCache for Memcached clusters. The cluster to be tagged is specified using an ARN (Amazon Resource Name).


**Topics**

- Listing Tags Using the ElastiCache API (p. 183)
- Adding Tags Using the ElastiCache API (p. 184)
- Modifying Tags Using the ElastiCache API (p. 184)
- Removing Tags Using the ElastiCache API (p. 184)

**Listing Tags Using the ElastiCache API**

You can use the ElastiCache API to list tags on an existing resource by using the `ListTagsForResource` operation.
Managing Tags Using the ElastiCache API

The following code uses the ElastiCache API to list the tags on the resource `my-cluster` in the us-west-2 region.

```plaintext
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Version=2015-02-02
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>
```

Adding Tags Using the ElastiCache API

You can use the ElastiCache API to add tags to an existing ElastiCache cluster by using the `AddTagsToResource` operation. If the tag key does not exist on the resource, the key and value are added to the resource. If the key already exists on the resource, the value associated with that key is updated to the new value.

The following code uses the ElastiCache API to add the keys `Service` and `Region` with the values `elasticache` and `us-west-2` respectively to the resource `my-cluster` in the us-west-2 region.

```plaintext
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Tags.member.1.Key=Service
&Tags.member.1.Value=elasticache
&Tags.member.2.Key=Region
&Tags.member.2.Value=us-west-2
&Version=2015-02-02
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>
```

For more information, see `AddTagsToResource` in the Amazon ElastiCache API Reference.

Modifying Tags Using the ElastiCache API

You can use the ElastiCache API to modify the tags on an ElastiCache cluster.

To modify the value of a tag:

- Use `AddTagsToResource` operation to either add a new tag and value or to change the value of an existing tag.
- Use `RemoveTagsFromResource` to remove tags from the resource.

Output from either operation will be a list of tags and their values on the specified resource.

Use `RemoveTagsFromResource` to remove tags from the resource.

Removing Tags Using the ElastiCache API

You can use the ElastiCache API to remove tags from an existing ElastiCache for Memcached cluster by using the `RemoveTagsFromResource` operation.

The following code uses the ElastiCache API to remove the tags with the keys `Service` and `Region` from the cluster `my-cluster` in region us-west-2.
Managing Tags Using the ElastiCache API

https://elasticache.us-west-2.amazonaws.com/
    ?Action=RemoveTagsFromResource
    &SignatureVersion=4
    &SignatureMethod=HmacSHA256
    &TagKeys.member.1=Service
    &TagKeys.member.2=Region
    &Version=2015-02-02
    &Timestamp=20150202T192317Z
    &X-Amz-Credential=<credential>
Managing Costs with Reserved Nodes

Reserving one or more nodes may be a way for you to reduce costs. Reserved nodes are charged an up front fee that depends upon the node type and the length of reservation—one or three years. In addition to the upfront charge, there is an hourly usage charge. This charge is much less than the hourly usage charge that you incur with On-Demand nodes.

To see if reserved nodes are a cost savings for your use cases, first determine the node size and number of nodes you need. Then estimate the usage of the node, and compare the total cost to you using On-Demand nodes versus reserved nodes. You can mix and match reserved and On-Demand node usage in your clusters. For pricing information, see Amazon ElastiCache Pricing.

You can use the AWS Management Console, the AWS CLI, or the ElastiCache API to list and purchase available reserved node offerings.

For more information on reserved nodes, see Amazon ElastiCache Reserved Cache Nodes.

Topics
- Standard Reserved Node Cache Offerings (p. 186)
- Legacy Reserved Node Cache Offerings (p. 186)
- Getting Info About Reserved Node Offerings (p. 189)
- Purchasing a Reserved Node (p. 192)
- Getting Info About Your Reserved Nodes (p. 195)

Standard Reserved Node Cache Offerings

When you purchase a standard reserved node instance (RI) in Amazon ElastiCache, you purchase a commitment to getting a discounted rate on a specific cache node instance type and AWS Region for the duration of the reserved node instance. To use an Amazon ElastiCache reserved node instance, you create a new ElastiCache node instance, just as you would for an on-demand instance.

The new node instance that you create must exactly match the specifications of the reserved node instance. If the specifications of the new node instance match an existing reserved node instance for your account, you are billed at the discounted rate offered for the reserved node instance. Otherwise, the node instance is billed at an on-demand rate. These standard RIs are available from R5 and M5 instance families onwards.

Note
All three offering types discussed next are available in one-year and three-year terms.

Offering Types

No Upfront RI provides access to a reserved ElastiCache instance without requiring an upfront payment. Your No Upfront reserved ElastiCache instance bills a discounted hourly rate for every hour within the term, regardless of usage.

Partial Upfront RI requires a part of the reserved Elasticache instance to be paid upfront. The remaining hours in the term are billed at a discounted hourly rate, regardless of usage. This option is the replacement for the legacy Heavy Utilization option, which is explained in the next section.

All Upfront RI requires full payment to be made at the start of the RI term. You incur no other costs for the remainder of the term, regardless of the number of hours used.

Legacy Reserved Node Cache Offerings

There are three levels of legacy node reservations—Heavy Utilization, Medium Utilization, and Light Utilization. Nodes can be reserved at any utilization level for either one or three years. The node type,
utilization level, and reservation term affect your total costs. Verify the savings that reserved nodes can provide your business by comparing various models before you purchase reserved nodes.

Nodes purchased at one utilization level or term cannot be converted to a different utilization level or term.

**Utilization Levels**

*Heavy Utilization reserved nodes* enable workloads that have a consistent baseline of capacity or run steady-state workloads. Heavy Utilization reserved nodes require a high up-front commitment, but if you plan to run more than 79 percent of the reserved node term you can earn the largest savings (up to 70 percent off of the On-Demand price). With Heavy Utilization reserved nodes, you pay a one-time fee. This is then followed by a lower hourly fee for the duration of the term regardless of whether your node is running.

*Medium Utilization reserved nodes* are the best option if you plan to use your reserved nodes a large amount of the time and you want either a lower one-time fee or to stop paying for your node when you shut it off. Medium Utilization reserved nodes are a more cost-effective option when you plan to run more than 40 percent of the reserved nodes term. This option can save you up to 64 percent off of the On-Demand price. With Medium Utilization reserved nodes, you pay a slightly higher one-time fee than with Light Utilization reserved nodes, and you receive lower hourly usage rates when you run a node.

*Light Utilization reserved nodes* are ideal for periodic workloads that run only a couple of hours a day or a few days per week. Using Light Utilization reserved nodes, you pay a one-time fee followed by a discounted hourly usage fee when your node is running. You can start saving when your node is running more than 17 percent of the reserved node term. You can save up to 56 percent off of the On-Demand rates over the entire term of your reserved node.

**Legacy Reserved Cache Node Offerings**

<table>
<thead>
<tr>
<th>Offering</th>
<th>Up-Front Cost</th>
<th>Usage Fee</th>
<th>Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy Utilization</td>
<td>Highest</td>
<td>Lowest</td>
<td>Hourly fee charged for each hour the node is running. No hourly charge when the node is not running. Lowest overall cost if you plan to run your reserved nodes more than 79 percent of a three-year term.</td>
</tr>
<tr>
<td>Medium Utilization</td>
<td>Medium</td>
<td>Hourly</td>
<td>Hourly usage fee charged for each hour the node is running. No hourly charge when the node is not running. Suitable for elastic workloads or when you expect moderate usage, more than 40 percent of a three-year term.</td>
</tr>
<tr>
<td>Light Utilization</td>
<td>Lowest</td>
<td>Hourly</td>
<td>Hourly usage fee charged for each hour the node is running. No hourly charge when the node is not running. Highest overall cost if you plan to run all of the time. However, this is the lowest overall cost if you plan to use your reserved node infrequently, more than about 15 percent of a three-year term.</td>
</tr>
<tr>
<td>On-Demand Use (No reserved nodes)</td>
<td>None</td>
<td>Highest</td>
<td>Hourly fee charged for each hour the node is running. Highest hourly cost.</td>
</tr>
</tbody>
</table>

API Version 2015-02-02

187
For more information, see Amazon ElastiCache Pricing.
Getting Info About Reserved Node Offerings

Before you purchase reserved nodes, you can get information about available reserved node offerings.

The following examples show how to get pricing and information about available reserved node offerings using the AWS Management Console, AWS CLI, and ElastiCache API.

**Topics**
- Getting Info About Reserved Node Offerings (Console) (p. 189)
- Getting Info About Reserved Node Offerings (AWS CLI) (p. 189)
- Getting Info About Reserved Node Offerings (ElastiCache API) (p. 190)

**Getting Info About Reserved Node Offerings (Console)**

To get pricing and other information about available reserved cluster offerings using the AWS Management Console, use the following procedure.

**To get information about available reserved node offerings**

2. In the navigation pane, choose **Reserved Cache Nodes**.
3. Choose **Purchase Reserved Cache Node**.
4. For **Product Description**, choose Memcached.
5. To determine the available offerings, make selections for the next three lists:
   - **Cache Node Type**
   - **Term**
   - **Offering Type**

   After you make these selections, the cost per node and total cost of your selections is shows in the **Purchase Reserved Cache Nodes** wizard.
6. Choose **Cancel** to avoid purchasing these nodes and incurring charges.

**Getting Info About Reserved Node Offerings (AWS CLI)**

To get pricing and other information about available reserved node offerings, type the following command at a command prompt:

```bash/aws elasticache describe-reserved-cache-nodes-offerings
```

This operation produces output similar to the following (JSON format):

```json
{
   "ReservedCacheNodesOfferings": [
   {
      "OfferingType": "Heavy Utilization",
      "FixedPrice": 4328.0,
      "ReservedCacheNodesOfferingId": "0192caa9-daf2-4159-b1e5-a79bb1916695",
      "UsagePrice": 0.0,
      "RecurringCharges": []
   }
```
Getting Info About Reserved Node Offerings (ElastiCache API)

To get pricing and information about available reserved node offerings, call the DescribeReservedCacheNodesOfferings action.

Example

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeReservedCacheNodesOfferings
&Version=2014-12-01
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&X-Amz-Algorithm
&X-Amz-SignedHeaders=Host
&X-Amz-Expires=20141201T220302Z
&X-Amz-Credential=<credential>
&X-Amz-Signature=<signature>
```

This call returns output similar to the following:

```
  <ReservedCacheNodesOfferingsResult>
    <ReservedCacheNodesOffering>
      <Duration>31536000</Duration>
      <OfferingType>Medium Utilization</OfferingType>
      <CurrencyCode>USD</CurrencyCode>
      <RecurringCharges/>
      <FixedPrice>1820.0</FixedPrice>
    </ReservedCacheNodesOffering>
  </ReservedCacheNodesOfferingsResult>
</DescribeReservedCacheNodesOfferingsResponse>
```
For more information, see DescribeReservedCacheNodesOfferings in the ElastiCache API Reference.
Purchasing a Reserved Node

The following examples show how to purchase a reserved node offering using the AWS Management Console, the AWS CLI, and the ElastiCache API.

**Important**
Following the examples in this section incurs charges on your AWS account that you can't reverse.

**Topics**
- Purchasing a Reserved Node (Console) (p. 192)
- Purchasing a Reserved Node (AWS CLI) (p. 192)
- Purchasing a Reserved Node (ElastiCache API) (p. 193)

### Purchasing a Reserved Node (Console)

This example shows purchasing a specific reserved node offering, 649fd0c8-cf6d-47a0-bfa6-060f8e75e95f, with a reserved node ID of `myreservationID`.

The following procedure uses the AWS Management Console to purchase the reserved node offering by offering id.

**To purchase reserved nodes**

2. In the navigation list, choose the **Reserved Cache Nodes** link.
3. Choose the **Purchase Reserved Cache Node** button.
4. Choose the node type from the **Product Description** drop-down list box.
5. Choose the node class from the **Cache Node Class** drop-down list box.
6. Choose length of time you want to reserve the node for from the **Term** drop-down list box.
7. Do either one of the following:
   - Choose the offering type from the **Offering Type** drop-down list box.
   - Enter a reserved node ID in the **Reserved Cache Node ID** text box.

   **Note**
   The Reserved Cache Node ID is a unique customer-specified identifier to track this reservation. If this box is left blank, ElastiCache automatically generates an identifier for the reservation.

8. Choose the **Next** button.

   The **Purchase Reserved Cache Node** dialog box shows a summary of the reserved node attributes that you've chosen and the payment due.

9. Choose the **Yes, Purchase** button to proceed and purchase the reserved node.

   **Important**
   When you choose **Yes, Purchase** you incur the charges for the reserved nodes you selected.
   To avoid incurring these charges, choose **Cancel**.

### Purchasing a Reserved Node (AWS CLI)

The following example shows purchasing the specific reserved cluster offering, 649fd0c8-cf6d-47a0-bfa6-060f8e75e95f, with a reserved node ID of `myreservationID`.
Type the following command at a command prompt:

For Linux, macOS, or Unix:

```bash
aws elasticache purchase-reserved-cache-nodes-offering
  --reserved-cache-nodes-offering-id 649fd0c8-cf6d-47a0-bfa6-060f8e75e95f
  --reserved-cache-node-id myreservationID
```

For Windows:

```bash
aws elasticache purchase-reserved-cache-nodes-offering
  --reserved-cache-nodes-offering-id 649fd0c8-cf6d-47a0-bfa6-060f8e75e95f
  --reserved-cache-node-id myreservationID
```

The command returns output similar to the following:

<table>
<thead>
<tr>
<th>RESERVATION</th>
<th>ReservationID</th>
<th>Class</th>
<th>Start Time</th>
<th>Duration</th>
<th>Fixed Price</th>
<th>Usage Price</th>
<th>Count</th>
<th>State</th>
<th>Description</th>
<th>Offering Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESERVATION</td>
<td>myreservationid</td>
<td>cache.m1.small</td>
<td>2013-12-19T00:30:23.247Z</td>
<td>1y</td>
<td>455.00 USD</td>
<td>0.092 USD</td>
<td>1</td>
<td>payment-pending</td>
<td>memcached</td>
<td>Medium Utilization</td>
</tr>
</tbody>
</table>

For more information, see `purchase-reserved-cache-nodes-offering` in the AWS CLI Reference.

### Purchasing a Reserved Node (ElastiCache API)

The following example shows purchasing the specific reserved node offering, 649fd0c8-cf6d-47a0-bfa6-060f8e75e95f, with a reserved cluster ID of `myreservationID`.

Call the `PurchaseReservedCacheNodesOffering` operation with the following parameters:

- `ReservedCacheNodesOfferingId` = 649fd0c8-cf6d-47a0-bfa6-060f8e75e95f
- `ReservedCacheNodeID` = myreservationID
- `CacheNodeCount` = 1

**Example**

```xml
https://elasticache.us-west-2.amazonaws.com/
  ?Action=PurchaseReservedCacheNodesOffering
  &ReservedCacheNodesOfferingId=649fd0c8-cf6d-47a0-bfa6-060f8e75e95f
  &ReservedCacheNodeID=myreservationID
  &CacheNodeCount=1
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20141201T220302Z
  &X-Amz-Algorithm=AWS4-HMAC-SHA256
  &X-Amz-Date=20141201T220302Z
  &X-Amz-SignedHeaders=Host
  &X-Amz-Expires=20141201T220302Z
  &X-Amz-Credential=<credential>
  &X-Amz-Signature=<signature>
```

This call returns output similar to the following:

```xml
  <PurchaseReservedCacheNodesOfferingResult>
    <ReservedCacheNode>
```

API Version 2015-02-02
For more information, see `PurchaseReservedCacheNodesOffering` in the ElastiCache API Reference.
Getting Info About Your Reserved Nodes

You can get information about the reserved nodes you've purchased using the AWS Management Console, the AWS CLI, and the ElastiCache API.

Topics

• Getting Info About Your Reserved Nodes (Console) (p. 195)
• Getting Info About Your Reserved Nodes (AWS CLI) (p. 195)
• Getting Info About Your Reserved Nodes (ElastiCache API) (p. 195)

Getting Info About Your Reserved Nodes (Console)

The following procedure describes how to use the AWS Management Console to get information about the reserved nodes you purchased.

To get information about your purchased reserved nodes

2. In the navigation list, choose the Reserved Cache Nodes link.

   The reserved nodes for your account appear in the Reserved Cache Nodes list. You can choose any of the reserved nodes in the list to see detailed information about the reserved node in the detail pane at the bottom of the console.

Getting Info About Your Reserved Nodes (AWS CLI)

To get information about reserved nodes for your AWS account, type the following command at a command prompt:

```
aws elasticache describe-reserved-cache-nodes
```

This operation produces output similar to the following (JSON format):

```
{
  "ReservedCacheNodeId": "myreservationid",
  "ReservedCacheNodesOfferingId": "649fd0c8-cf6d-47a0-bfa6-060f8e75e95f",
  "CacheNodeType": "cache.m1.small",
  "Duration": "31536000",
  "ProductDescription": "memcached",
  "OfferingType": "Medium Utilization",
  "MaxRecords": 0
}
```

For more information, see describe--reserved-cache-nodes in the AWS CLI Reference.

Getting Info About Your Reserved Nodes (ElastiCache API)

To get information about reserved nodes for your AWS account, call the DescribeReservedCacheNodes operation.

Example

```
https://elasticache.us-west-2.amazonaws.com/
```
Getting Info About Your Reserved Nodes

This call returns output similar to the following:

```
  <DescribeReservedCacheNodesResult>
    <ReservedCacheNodes>
      <ReservedCacheNode>
        <OfferingType>Medium Utilization</OfferingType>
        <CurrencyCode>USD</CurrencyCode>
        <RecurringCharges/>
        <ProductDescription>memcached</ProductDescription>
        <ReservedCacheNodesOfferingId>649fd0c8-cf6d-47a0-bfa6-060f8e75e95f</ReservedCacheNodesOfferingId>
        <State>payment-failed</State>
        <ReservedCacheNodeId>myreservationid</ReservedCacheNodeId>
        <CacheNodeCount>1</CacheNodeCount>
        <StartTime>2010-12-15T00:25:14.131Z</StartTime>
        <Duration>31536000</Duration>
        <FixedPrice>227.5</FixedPrice>
        <UsagePrice>0.046</UsagePrice>
        <CacheNodeType>cache.m1.small</CacheNodeType>
      </ReservedCacheNode>
      (...) some output omitted for brevity...
    </ReservedCacheNodes>
  </DescribeReservedCacheNodesResult>
  <ResponseMetadata>
    <RequestId>23400d50-2978-11e1-9e6d-771388d6ed6b</RequestId>
  </ResponseMetadata>
</DescribeReservedCacheNodesResponse>

For more information, see DescribeReservedCacheNodes in the ElastiCache API Reference.
Security in Amazon ElastiCache

Cloud security at AWS is the highest priority. As an AWS customer, you benefit from a data center and network architecture that is built to meet the requirements of the most security-sensitive organizations.

Security is a shared responsibility between AWS and you. The shared responsibility model describes this as security of the cloud and security in the cloud:

- **Security of the cloud** – AWS is responsible for protecting the infrastructure that runs AWS services in the AWS Cloud. AWS also provides you with services that you can use securely. Third-party auditors regularly test and verify the effectiveness of our security as part of the AWS compliance programs. To learn about the compliance programs that apply to Amazon ElastiCache, see AWS Services in Scope by Compliance Program.
- **Security in the cloud** – Your responsibility is determined by the AWS service that you use. You are also responsible for other factors including the sensitivity of your data, your company’s requirements, and applicable laws and regulations.

This documentation helps you understand how to apply the shared responsibility model when using Amazon ElastiCache. The following topics show you how to configure Amazon ElastiCache to meet your security and compliance objectives. You also learn how to use other AWS services that help you to monitor and secure your Amazon ElastiCache resources.

**Topics**
- Data Protection in Amazon ElastiCache (p. 197)
- Internetwork Traffic Privacy (p. 198)
- Identity and Access Management in Amazon ElastiCache (p. 230)
- Compliance Validation for Amazon ElastiCache (p. 251)
- Resilience in Amazon ElastiCache (p. 251)
- Infrastructure Security in AWS Elastiecache (p. 253)

Data Protection in Amazon ElastiCache

Amazon ElastiCache conforms to the AWS shared responsibility model, which includes regulations and guidelines for data protection. AWS is responsible for protecting the global infrastructure that runs all the AWS services. AWS maintains control over data hosted on this infrastructure, including the security configuration controls for handling customer content and personal data. AWS customers and APN partners, acting either as data controllers or data processors, are responsible for any personal data that they put in the AWS Cloud.

For data protection purposes, we recommend that you protect AWS account credentials and set up individual user accounts with AWS Identity and Access Management (IAM). In this approach, each user is given only the permissions necessary to fulfill their job duties. We also recommend that you secure your data in the following ways:

- Use multi-factor authentication (MFA) with each account.
- Use SSL/TLS to communicate with AWS resources.
- Set up API and user activity logging with AWS CloudTrail.
- Use AWS encryption solutions, along with all default security controls within AWS services.
• Use advanced managed security services such as Amazon Macie, which assists in discovering and securing personal data that is stored in Amazon S3.

We strongly recommend that you never put sensitive identifying information, such as your customers' account numbers, into free-form fields such as a Name field. This includes when you work with Amazon ElastiCache or other AWS services using the console, API, AWS CLI, or AWS SDKs. Any data that you enter into Amazon ElastiCache or other services might get picked up for inclusion in diagnostic logs. When you provide a URL to an external server, don't include credentials information in the URL to validate your request to that server.

For more information about data protection, see the AWS Shared Responsibility Model and GDPR blog post on the AWS Security Blog.

Internetwork Traffic Privacy

Amazon ElastiCache uses the following techniques to secure your cache data and protect it from unauthorized access:

• Amazon VPCs and ElastiCache Security (p. 198) explains the type of security group you need for your installation.
• Identity and Access Management in Amazon ElastiCache (p. 230) for granting and limiting actions of users, groups, and roles.

Amazon VPCs and ElastiCache Security

Because data security is important, ElastiCache provides means for you to control who has access to your data. How you control access to your data is dependent upon whether or not you launched your clusters in an Amazon Virtual Private Cloud (Amazon VPC) or Amazon EC2-Classic.

Important
We have deprecated the use of Amazon EC2-Classic for launching ElastiCache clusters. All current generation nodes are launched in Amazon Virtual Private Cloud only.

Topics
• Understanding ElastiCache and Amazon VPCs (p. 200)
• Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 204)
• Creating a Virtual Private Cloud (VPC) (p. 210)
• Creating a Cache Subnet Group (p. 212)
• Creating a Cache Cluster in an Amazon VPC (p. 213)
• Connecting to a Cluster Running in an Amazon VPC (p. 214)

The Amazon Virtual Private Cloud (Amazon VPC) service defines a virtual network that closely resembles a traditional data center. When you configure your Amazon VPC you can select its IP address range, create subnets, and configure route tables, network gateways, and security settings. You can also add a cache cluster to the virtual network, and control access to the cache cluster by using Amazon VPC security groups.

This section explains how to manually configure an ElastiCache cluster in an Amazon VPC. This information is intended for users who want a deeper understanding of how ElastiCache and Amazon VPC work together.

Topics
• Understanding ElastiCache and Amazon VPCs (p. 200)
• Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 204)
• Creating a Virtual Private Cloud (VPC) (p. 210)
• Creating a Cache Subnet Group (p. 212)
• Creating a Cache Cluster in an Amazon VPC (p. 213)
• Connecting to a Cluster Running in an Amazon VPC (p. 214)
Understanding ElastiCache and Amazon VPCs

ElastiCache is fully integrated with the Amazon Virtual Private Cloud (Amazon VPC). For ElastiCache users, this means the following:

- If your AWS account supports only the EC2-VPC platform, ElastiCache always launches your cluster in an Amazon VPC.
- If you’re new to AWS, your clusters will be deployed into an Amazon VPC. A default VPC will be created for you automatically.
- If you have a default VPC and don’t specify a subnet when you launch a cluster, the cluster launches into your default Amazon VPC.

For more information, see Detecting Your Supported Platforms and Whether You Have a Default VPC.

With Amazon Virtual Private Cloud, you can create a virtual network in the AWS cloud that closely resembles a traditional data center. You can configure your Amazon VPC, including selecting its IP address range, creating subnets, and configuring route tables, network gateways, and security settings.

The basic functionality of ElastiCache is the same in a virtual private cloud; ElastiCache manages software upgrades, patching, failure detection and recovery whether your clusters are deployed inside or outside an Amazon VPC.

ElastiCache cache nodes deployed outside an Amazon VPC are assigned an IP address to which the endpoint/DNS name resolves. This provides connectivity from Amazon Elastic Compute Cloud (Amazon EC2) instances. When you launch an ElastiCache cluster into an Amazon VPC private subnet, every cache node is assigned a private IP address within that subnet.

Overview of ElastiCache In an Amazon VPC

The following diagram and table describe the Amazon VPC environment, along with ElastiCache clusters and Amazon EC2 instances that are launched in the Amazon VPC.
The Amazon VPC is an isolated portion of the AWS Cloud that is assigned its own block of IP addresses.

An Internet gateway connects your Amazon VPC directly to the Internet and provides access to other AWS resources such as Amazon Simple Storage Service (Amazon S3) that are running outside your Amazon VPC.

An Amazon VPC subnet is a segment of the IP address range of an Amazon VPC where you can isolate AWS resources according to your security and operational needs.

A routing table in the Amazon VPC directs network traffic between the subnet and the Internet. The Amazon VPC has an implied router, which is symbolized in this diagram by the circle with the R.
An Amazon VPC security group controls inbound and outbound traffic for your ElastiCache clusters and Amazon EC2 instances.

You can launch an ElastiCache cluster in the subnet. The cache nodes have private IP addresses from the subnet's range of addresses.

You can also launch Amazon EC2 instances in the subnet. Each Amazon EC2 instance has a private IP address from the subnet's range of addresses. The Amazon EC2 instance can connect to any cache node in the same subnet.

For an Amazon EC2 instance in your Amazon VPC to be reachable from the Internet, you need to assign a static, public address called an Elastic IP address to the instance.

Why use the Amazon VPC instead of EC2 Classic with your ElastiCache deployment?

Launching your instances into an Amazon VPC allows you to:

- Assign static private IP addresses to your instances that persist across starts and stops.
- Assign multiple IP addresses to your instances.
- Define network interfaces, and attach one or more network interfaces to your instances.
- Change security group membership for your instances while they’re running.
- Control the outbound traffic from your instances (egress filtering) in addition to controlling the inbound traffic to them (ingress filtering).
- Add an additional layer of access control to your instances in the form of network access control lists (ACL).
- Run your instances on single-tenant hardware.

For a comparison of Amazon EC2 Classic, Default VPC, and Non-default VPC, see Differences Between EC2-Classic and EC2-VPC.

The Amazon VPC must allow non-dedicated Amazon EC2 instances. You cannot use ElastiCache in an Amazon VPC that is configured for dedicated instance tenancy.

Prerequisites

To create an ElastiCache cluster within an Amazon VPC, your Amazon VPC must meet the following requirements:

- The Amazon VPC must allow nondedicated Amazon EC2 instances. You cannot use ElastiCache in an Amazon VPC that is configured for dedicated instance tenancy.
- A cache subnet group must be defined for your Amazon VPC. ElastiCache uses that cache subnet group to select a subnet and IP addresses within that subnet to associate with your cache nodes.
- A cache security group must be defined for your Amazon VPC, or you can use the default provided.
- CIDR blocks for each subnet must be large enough to provide spare IP addresses for ElastiCache to use during maintenance activities.

Routing and Security

You can configure routing in your Amazon VPC to control where traffic flows (for example, to the Internet gateway or virtual private gateway). With an Internet gateway, your Amazon VPC has direct access to other AWS resources that are not running in your Amazon VPC. If you choose to have only a virtual private gateway with a connection to your organization's local network, you can route your...
Internet-bound traffic over the VPN and use local security policies and firewall to control egress. In that case, you incur additional bandwidth charges when you access AWS resources over the Internet.

You can use Amazon VPC security groups to help secure the ElastiCache clusters and Amazon EC2 instances in your Amazon VPC. Security groups act like a firewall at the instance level, not the subnet level.

**Note**
We strongly recommend that you use DNS names to connect to your cache nodes, as the underlying IP address can change if you reboot the cache node.

### Amazon VPC Documentation

Amazon VPC has its own set of documentation to describe how to create and use your Amazon VPC. The following table gives links to the Amazon VPC guides.

<table>
<thead>
<tr>
<th>Description</th>
<th>Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>How to get started using Amazon VPC</td>
<td>Amazon VPC Getting Started Guide</td>
</tr>
<tr>
<td>How to use Amazon VPC through the AWS Management Console</td>
<td>Amazon VPC User Guide</td>
</tr>
<tr>
<td>Complete descriptions of all the Amazon VPC commands</td>
<td>Amazon EC2 Command Line Reference (the Amazon VPC commands are part of the Amazon EC2 reference)</td>
</tr>
<tr>
<td>Complete descriptions of the Amazon VPC API actions, data types, and errors</td>
<td>Amazon EC2 API Reference (the Amazon VPC API actions are part of the Amazon EC2 reference)</td>
</tr>
<tr>
<td>Information for the network administrator who needs to configure the gateway at your end of an optional IPsec VPN connection</td>
<td>AWS Site-to-Site VPN Network Administrator Guide</td>
</tr>
</tbody>
</table>

For more detailed information about Amazon Virtual Private Cloud, see [Amazon Virtual Private Cloud](#).
Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC

Amazon ElastiCache supports the following scenarios for accessing a cluster in an Amazon VPC:

Contents

- Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in the Same Amazon VPC (p. 204)
- Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs (p. 205)
  - Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in the Same Region (p. 206)
  - Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in Different Regions (p. 207)
- Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center (p. 207)
  - Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using VPN Connectivity (p. 208)
  - Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using Direct Connect (p. 209)

Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in the Same Amazon VPC

The most common use case is when an application deployed on an EC2 instance needs to connect to a cluster in the same VPC.

The following diagram illustrates this scenario

The simplest way to manage access between EC2 instances and DB instances in the same VPC is to do the following:

1. Create a VPC security group for your cluster. This security group can be used to restrict access to the cluster instances. For example, you can create a custom rule for this security group that allows TCP access using the port you assigned to the cluster when you created it and an IP address you will use to access the cluster.

   The default port for Memcached clusters is 11211.

2. Create a VPC security group for your EC2 instances (web and application servers). This security group can, if needed, allow access to the EC2 instance from the Internet via the VPC's routing table. For
example, you can set rules on this security group to allow TCP access to the EC2 instance over port 22.

3. Create custom rules in the security group for your Cluster that allow connections from the security group you created for your EC2 instances. This would allow any member of the security group to access the DB instances.

To create a rule in a VPC security group that allows connections from another security group

1. Sign in to the AWS Management Console and open the Amazon VPC console at https://console.aws.amazon.com/vpc.

2. In the navigation pane, choose Security Groups.

3. Select or create a security group that you will use for your Cluster instances. Choose Add Rule. This security group will allow access to members of another security group.

4. From Type choose Custom TCP Rule.
   a. For Port Range, specify the port you used when you created your cluster.
      The default port for Memcached clusters is 11211.
   b. In the Source box, start typing the ID of the security group. From the list select the security group you will use for your Amazon EC2 instances.

5. Choose Save when you finish.

Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs

When your Cluster is in a different VPC from the EC2 instance you are using to access it, there are several ways to access the DB instance. If the Cluster and EC2 instance are in different VPCs but in the same region, you can use VPC peering. If the Cluster and the EC2 instance are in different regions, you can create VPN connectivity between regions.

Topics
- Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in the Same Region (p. 206)
- Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in Different Regions (p. 207)
Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in the Same Region

The following diagram illustrates accessing a cluster by an Amazon EC2 instance in a different Amazon VPC in the same region using an Amazon VPC peering connection.

Cluster accessed by an Amazon EC2 instance in a different Amazon VPC within the same Region - VPC Peering Connection

A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IP addresses. Instances in either VPC can communicate with each other as if they are within the same network. You can create a VPC peering connection between your own Amazon VPCs, or with an Amazon VPC in another AWS account within a single region. To learn more about Amazon VPC peering, see the VPC documentation.

To access a cluster in a different Amazon VPC over peering

1. Make sure that the two VPCs do not have an overlapping IP range or you will not be able to peer them.
2. Peer the two VPCs. For more information, see Creating and Accepting an Amazon VPC Peering Connection.
3. Update your routing table. For more information, see Updating Your Route Tables for a VPC Peering Connection

Following is what the route tables look like for the example in the preceding diagram. Note that pcx-a894f1c1 is the peering connection.

<table>
<thead>
<tr>
<th>Destination</th>
<th>Target</th>
<th>Destination</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>172.16.0.0/16</td>
<td>local</td>
<td>10.10.0.0/16</td>
<td>local</td>
</tr>
<tr>
<td>10.10.0.0/16</td>
<td>pcx-a894f1c1</td>
<td>0.0.0.0/0</td>
<td>igw-bf1ccd8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>172.16.0.0/16</td>
<td>pcx-a894f1c1</td>
</tr>
</tbody>
</table>

VPC Routing Table

4. Modify the Security Group of your ElastiCache cluster to allow inbound connection from the Application security group in the peered VPC. For more information, see Reference Peer VPC Security Groups.

Accessing a cluster over a peering connection will incur additional data transfer costs.
Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in Different Regions

One common strategy for connecting multiple, geographically disperse VPCs and remote networks is to create a transit VPC that serves as a global network transit center. A transit VPC simplifies network management and minimizes the number of connections required to connect multiple VPCs and remote networks. This design can save time and effort and also reduce costs, as it is implemented virtually without the traditional expense of establishing a physical presence in a colocation transit hub or deploying physical network gear.

Connecting across different VPCs in different regions

Once the Transit Amazon VPC is established, an application deployed in a “spoke” VPC in one region can connect to an ElastiCache cluster in a “spoke” VPC within another region.

To access a cluster in a different VPC within a different Region

1. Deploy a Transit VPC Solution. For more information, see, How do I build a global transit network on AWS?.
2. Update the VPC routing tables in the App and Cache VPCs to route traffic through the VGW (Virtual Private Gateway) and the VPN Appliance. In case of Dynamic Routing with Border Gateway Protocol (BGP) your routes may be automatically propagated.
3. Modify the Security Group of your ElastiCache cluster to allow inbound connection from the Application instances IP range. Note that you will not be able to reference the application server Security Group in this scenario.

Accessing a cluster across regions will introduce networking latencies and additional cross-region data transfer costs.

Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center

Another possible scenario is a Hybrid architecture where clients or applications in the customer's data center may need to access an ElastiCache Cluster in the VPC. This scenario is also supported providing there is connectivity between the customers' VPC and the data center either through VPN or Direct Connect.
Topics

- Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using VPN Connectivity (p. 208)
- Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using Direct Connect (p. 209)

Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using VPN Connectivity

The following diagram illustrates accessing an ElastiCache cluster from an application running in your corporate network using VPN connections.

Connecting to ElastiCache from your data center via a VPN

To access a cluster in a VPC from on-prem application over VPN connection

1. Establish VPN Connectivity by adding a hardware Virtual Private Gateway to your VPC. For more information, see Adding a Hardware Virtual Private Gateway to Your VPC.
2. Update the VPC routing table for the subnet where your ElastiCache cluster is deployed to allow traffic from your on-premises application server. In case of Dynamic Routing with BGP your routes may be automatically propagated.
3. Modify the Security Group of your ElastiCache cluster to allow inbound connection from the on-premises application servers.
Accessing a cluster over a VPN connection will introduce networking latencies and additional data transfer costs.

**Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using Direct Connect**

The following diagram illustrates accessing an ElastiCache cluster from an application running on your corporate network using Direct Connect.

**Connecting to ElastiCache from your data center via Direct Connect**

**To access an ElastiCache cluster from an application running in your network using Direct Connect**

1. Establish Direct Connect connectivity. For more information, see, *Getting Started with AWS Direct Connect*.
2. Modify the Security Group of your ElastiCache cluster to allow inbound connection from the on-premises application servers.

Accessing a cluster over DX connection may introduce networking latencies and additional data transfer charges.
Creating a Virtual Private Cloud (VPC)

In this example, you create an Amazon VPC with a private subnet for each Availability Zone.

Creating an Amazon VPC (Console)

To create an ElastiCache cluster inside an Amazon Virtual Private Cloud

1. Sign in to the AWS Management Console, and open the Amazon VPC console at https://console.aws.amazon.com/vpc/.
2. Create a new Amazon VPC by using the Amazon Virtual Private Cloud wizard:
   a. In the navigation list, choose VPC Dashboard.
   b. Choose Start VPC Wizard.
   c. In the Amazon VPC wizard, choose VPC with Public and Private Subnets, and then choose Next.
   d. On the VPC with Public and Private Subnets page, keep the default options, and then choose Create VPC.
   e. In the confirmation message that appears, choose Close.
3. Confirm that there are two subnets in your Amazon VPC, a public subnet and a private subnet. These subnets are created automatically.
   a. In the navigation list, choose Subnets.
   b. In the list of subnets, find the two subnets that are in your Amazon VPC:

   ![Subnet ID State VPC ID CIDR Block Available IPs]
   - subnet-5bf5e639: available, vpc-d3a77cb6: 10.0.1.0/24: 256: available
   - subnet-5bf5e63a: available, vpc-d3a77cb6: 10.0.0.0/24: 250: available

   The public subnet will have one fewer available IP address, because the wizard creates an Amazon EC2 NAT instance and an Elastic IP address (for which Amazon EC2 rates apply) for outbound communication to the Internet from your private subnet.

   **Tip**
   Make a note of your two subnet identifiers, and which is public and private. You will need this information later when you launch your cache clusters and add an Amazon EC2 instance to your Amazon VPC.
4. Create an Amazon VPC security group. You will use this group for your cache cluster and your Amazon EC2 instance.
   a. In the navigation pane of the Amazon VPC Management console, choose Security Groups.
   c. Type a name and a description for your security group in the corresponding boxes. In the VPC box, choose the identifier for your Amazon VPC.
When the settings are as you want them, choose Yes, Create.

5. Define a network ingress rule for your security group. This rule will allow you to connect to your Amazon EC2 instance using Secure Shell (SSH).
   a. In the navigation list, choose Security Groups.
   b. Find your security group in the list, and then choose it.
   c. Under Security Group, choose the Inbound tab. In the Create a new rule box, choose SSH, and then choose Add Rule.
   d. Choose Apply Rule Changes.

Now you are ready to create a cache subnet group and launch a cache cluster in your Amazon VPC.
Creating a Cache Subnet Group

A cache subnet group is a collection of subnets that you may want to designate for your cache clusters in an Amazon VPC. When launching a cache cluster in an Amazon VPC, you need to select a cache subnet group. Then ElastiCache uses that cache subnet group to assign IP addresses within that subnet to each cache node in the cluster.

For guidance on how to create a subnet group using the ElastiCache Management Console, the AWS CLI, or the ElastiCache API, see Creating a Subnet Group (p. 215).

After you create a cache subnet group, you can launch a cache cluster to run in your Amazon VPC. Continue to the next topic Creating a Cache Cluster in an Amazon VPC (p. 213).
Creating a Cache Cluster in an Amazon VPC

In this example, you create a cache cluster in your Amazon VPC.

Creating a Cache Cluster in an Amazon VPC (Console)

To launch a Memcached cache cluster, see Creating a Memcached Cluster (Console) (p. 82). In step 6.c select a VPC subnet group.

You have now launched a cache cluster inside an Amazon VPC. For an example of one way to connect to your new cache cluster running in the Amazon VPC, continue to Connecting to a Cluster Running in an Amazon VPC (p. 214).
Connecting to a Cluster Running in an Amazon VPC

This example shows how to launch an Amazon EC2 instance in your Amazon VPC. You can then log in to this instance and access the ElastiCache cluster that is running in the Amazon VPC.

**Note**
For information about using Amazon EC2, see the Amazon EC2 Getting Started Guide in the Amazon EC2 documentation.

**Important**
To avoid incurring additional charges on your AWS account, be sure to delete any AWS resources you no longer want after trying these examples.

For information on connecting to your cluster, see Step 3: Connect to a Cluster's Node (p. 20) in the ElastiCache User Guide.

Subnets and Subnet Groups

A *subnet group* is a collection of subnets (typically private) that you can designate for your clusters running in an Amazon Virtual Private Cloud (VPC) environment.

If you create a cluster in an Amazon VPC, you must specify a subnet group. ElastiCache uses that subnet group to choose a subnet and IP addresses within that subnet to associate with your nodes.

This section covers how to create and leverage subnets and subnet groups to manage access to your ElastiCache resources.

For more information about subnet group usage in an Amazon VPC environment, see Step 2: Authorize Access (p. 19).

**Topics**
- Creating a Subnet Group (p. 215)
- Assigning a Subnet Group to a Cluster (p. 218)
- Modifying a Subnet Group (p. 219)
- Deleting a Subnet Group (p. 221)
Creating a Subnet Group

When you create a new subnet group, note the number of available IP addresses. If the subnet has very few free IP addresses, you might be constrained as to how many more nodes you can add to the cluster. To resolve this issue, you can assign one or more subnets to a subnet group so that you have a sufficient number of IP addresses in your cluster's Availability Zone. After that, you can add more nodes to your cluster.

The following procedures show you how to create a subnet group called mysubnetgroup (console), the AWS CLI, and the ElastiCache API.

Creating a Subnet Group (Console)

The following procedure shows how to create a subnet group (console).

To create a subnet group (Console)

2. In the navigation list, choose Subnet Groups.
3. Choose Create Subnet Group.
4. In the Create Subnet Group wizard, do the following. When all the settings are as you want them, choose Yes, Create.
   a. In the Name box, type a name for your subnet group.
   b. In the Description box, type a description for your subnet group.
   c. In the VPC ID box, choose the Amazon VPC that you created.
   d. In the Availability Zone and Subnet ID lists, choose the Availability Zone and ID of your private subnet, and then choose Add.

5. In the confirmation message that appears, choose Close.
Your new subnet group appears in the **Subnet Groups** list of the ElastiCache console. At the bottom of the window you can choose the subnet group to see details, such as all of the subnets associated with this group.

### Creating a Subnet Group (AWS CLI)

At a command prompt, use the command `create-cache-subnet-group` to create a subnet group.

For Linux, macOS, or Unix:

```bash
aws elasticache create-cache-subnet-group
--cache-subnet-group-name mysubnetgroup
--cache-subnet-group-description "Testing"
--subnet-ids subnet-53df9c3a
```

For Windows:

```bash
aws elasticache create-cache-subnet-group ^
--cache-subnet-group-name mysubnetgroup ^
--cache-subnet-group-description "Testing" ^
--subnet-ids subnet-53df9c3a
```

This command should produce output similar to the following:

```json
{
  "CacheSubnetGroup": {
    "VpcId": "vpc-37c3cd17",
    "CacheSubnetGroupDescription": "Testing",
    "Subnets": [
      {
        "SubnetIdentifier": "subnet-53df9c3a",
        "SubnetAvailabilityZone": {
          "Name": "us-west-2a"
        }
      }
    ],
    "CacheSubnetGroupName": "mysubnetgroup"
  }
}
```

For more information, see the AWS CLI topic [create-cache-subnet-group](https://docs.aws.amazon.com/cli/latest/reference/elasticache/create-cache-subnet-group.html).

### Creating a Subnet Group (ElastiCache API)

Using the ElastiCache API, call `CreateCacheSubnetGroup` with the following parameters:

- `CacheSubnetGroupName=mysubnetgroup`
- `CacheSubnetGroupDescription=Testing`
- `SubnetIds.member.1=subnet-53df9c3a`

**Example**

```xml
https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateCacheSubnetGroup
  &CacheSubnetGroupDescription=Testing
  &CacheSubnetGroupName=mysubnetgroup
  &SignatureMethod=HmacSHA256
  &SignatureVersion=4
```

API Version 2015-02-02

216
<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;SubnetIds.member.1</td>
<td>subnet-53df9c3a</td>
</tr>
<tr>
<td>&amp;Timestamp</td>
<td>20141201T220302Z</td>
</tr>
<tr>
<td>&amp;Version</td>
<td>2014-12-01</td>
</tr>
<tr>
<td>&amp;X-Amz-Algorithm</td>
<td>AWS4-HMAC-SHA256</td>
</tr>
<tr>
<td>&amp;X-Amz-Credential</td>
<td>&lt;credential&gt;</td>
</tr>
<tr>
<td>&amp;X-Amz-Date</td>
<td>20141201T220302Z</td>
</tr>
<tr>
<td>&amp;X-Amz-Expires</td>
<td>20141201T220302Z</td>
</tr>
<tr>
<td>&amp;X-Amz-Signature</td>
<td>&lt;signature&gt;</td>
</tr>
<tr>
<td>&amp;X-Amz-SignedHeaders</td>
<td>Host</td>
</tr>
</tbody>
</table>

API Version 2015-02-02
Assigning a Subnet Group to a Cluster

After you have created a subnet group, you can launch a cluster in an Amazon VPC. For more information, see the following.

- **Memcached cluster** – To launch a Memcached cluster, see Creating a Memcached Cluster (Console) (p. 82). In step 5.a (Advanced Memcached Settings), choose a VPC subnet group.
Modifying a Subnet Group

You can modify a subnet group's description, or modify the list of subnet IDs associated with the subnet group. You cannot delete a subnet ID from a subnet group if a cluster is currently using that subnet.

The following procedures show you how to modify a subnet group.

Modifying Subnet Groups (Console)

To modify a subnet group

2. In the navigation pane, choose Subnet Groups.
3. In the list of subnet groups, choose the one you want to modify.
4. In the lower portion of the ElastiCache console, make any changes to the description or the list of subnet IDs for the subnet group. To save your changes, choose Save.

Modifying Subnet Groups (AWS CLI)

At a command prompt, use the command modify-cache-subnet-group to modify a subnet group.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-cache-subnet-group \
    --cache-subnet-group-name mysubnetgroup \
    --cache-subnet-group-description "New description" \
    --subnet-ids "subnet-42df9c3a" "subnet-48fc21a9"
```

For Windows:

```bash
aws elasticache modify-cache-subnet-group ^
    --cache-subnet-group-name mysubnetgroup ^
    --cache-subnet-group-description "New description" ^
    --subnet-ids "subnet-42df9c3a" "subnet-48fc21a9"
```

This command should produce output similar to the following:

```json
{
    "CacheSubnetGroup": {
        "VpcId": "vpc-73cd3c17",
        "CacheSubnetGroupDescription": "New description",
        "Subnets": [
            {
                "SubnetIdentifier": "subnet-42df9c3a",
                "SubnetAvailabilityZone": {
                    "Name": "us-west-2a"
                }
            },
            {
                "SubnetIdentifier": "subnet-48fc21a9",
                "SubnetAvailabilityZone": {
                    "Name": "us-west-2a"
                }
            }
        ],
        "CacheSubnetGroupName": "mysubnetgroup"
    }
}
```
For more information, see the AWS CLI topic `modify-cache-subnet-group`.

**Modifying Subnet Groups (ElastiCache API)**

Using the ElastiCache API, call `ModifyCacheSubnetGroup` with the following parameters:

- `CacheSubnetGroupName=mysubnetgroup`
- Any other parameters whose values you want to change. This example uses `CacheSubnetGroupDescription=New%20description` to change the description of the subnet group.

**Example**

```plaintext
https://elasticache.us-west-2.amazonaws.com/
   ?Action=ModifyCacheSubnetGroup
   &CacheSubnetGroupDescription=New%20description
   &CacheSubnetGroupName=mysubnetgroup
   &SubnetIds.member.1=subnet-42df9c3a
   &SubnetIds.member.2=subnet-48fc21a9
   &SignatureMethod=HmacSHA256
   &SignatureVersion=4
   &Timestamp=20141201T220302Z
   &Version=2014-12-01
   &X-Amz-Algorithm=AWS4-HMAC-SHA256
   &X-Amz-Credential=<credential>
   &X-Amz-Date=20141201T220302Z
   &X-Amz-Expires=20141201T220302Z
   &X-Amz-Signature=<signature>
   &X-Amz-SignedHeaders=Host
```

**Note**

When you create a new subnet group, take note the number of available IP addresses. If the subnet has very few free IP addresses, you might be constrained as to how many more nodes you can add to the cluster. To resolve this issue, you can assign one or more subnets to a subnet group so that you have a sufficient number of IP addresses in your cluster's Availability Zone. After that, you can add more nodes to your cluster.
Deleting a Subnet Group

If you decide that you no longer need your subnet group, you can delete it. You cannot delete a subnet group if it is currently in use by a cluster.

The following procedures show you how to delete a subnet group.

Deleting a Subnet Group (Console)

To delete a subnet group

2. In the navigation pane, choose Subnet Groups.
3. In the list of subnet groups, choose the one you want to delete and then choose Delete.
4. When you are asked to confirm this operation, choose Yes, Delete.

Deleting a Subnet Group (AWS CLI)

Using the AWS CLI, call the command delete-cache-subnet-group with the following parameter:

• --cache-subnet-group-name mysubnetgroup

For Linux, macOS, or Unix:

aws elasticache delete-cache-subnet-group --cache-subnet-group-name mysubnetgroup

For Windows:

aws elasticache delete-cache-subnet-group --cache-subnet-group-name mysubnetgroup

This command produces no output.

For more information, see the AWS CLI topic delete-cache-subnet-group.

Deleting a Subnet Group (ElastiCache API)

Using the ElastiCache API, call DeleteCacheSubnetGroup with the following parameter:

• CacheSubnetGroupName=mysubnetgroup

Example

Line breaks are added for ease of reading.


API Version 2015-02-02

221
Important
Amazon ElastiCache security groups are only applicable to clusters that are not running in an Amazon Virtual Private Cloud environment (VPC). If you are running in an Amazon Virtual Private Cloud, **Security Groups** is not available in the console navigation pane.

If you are running your ElastiCache nodes in an Amazon VPC, you control access to your clusters with Amazon VPC security groups, which are different from ElastiCache security groups. For more information about using ElastiCache in an Amazon VPC, see Amazon VPCs and ElastiCache Security (p. 198).

Amazon ElastiCache allows you to control access to your clusters using ElastiCache security groups. An ElastiCache security group acts like a firewall, controlling network access to your cluster. By default, network access is turned off to your clusters. If you want your applications to access your cluster, you must explicitly enable access from hosts in specific Amazon EC2 security groups. Once ingress rules are configured, the same rules apply to all clusters associated with that security group.

To allow network access to your cluster, create a security group and use the `AuthorizeCacheSecurityGroupIngress` API operation (CLI: `authorize-cache-security-group-ingress`) to authorize the desired Amazon EC2 security group (which in turn specifies the Amazon EC2 instances allowed). The security group can be associated with your cluster at the time of creation, or using the `ModifyCacheCluster` API operation (CLI: `modify-cache-cluster`).

**Important**
Access control based on IP range is currently not enabled at the individual cluster level. All clients to a cluster must be within the EC2 network, and authorized via security groups as described previously.

For more information about using ElastiCache with Amazon VPCs, see Amazon VPCs and ElastiCache Security (p. 198).

Note that Amazon EC2 instances running in an Amazon VPC can't connect to ElastiCache clusters in EC2-Classic.

**Topics**
- Creating a Security Group (p. 223)
- Listing Available Security Groups (p. 225)
- Viewing a Security Group (p. 227)
- Authorizing Network Access to an Amazon EC2 Security Group (p. 229)
Creating a Security Group

This topic is relevant to you only if you are not running in an Amazon VPC. If you are running in an Amazon VPC, see Amazon VPCs and ElastiCache Security (p. 198).

To create a security group, you need to provide a name and a description.

The following procedures show you how to create a new security group.

Creating a Security Group (Console)

2. In the navigation pane, choose Security Groups.
4. In Create Security Group, type the name of the new security group in Security Group.
5. In Description, type a description for the new security group.
6. Choose Create.

Creating a Security Group (AWS CLI)

At a command prompt, use the create-cache-security-group command with the following parameters:

- --cache-security-group-name – The name of the security group you are creating.
  Example: mysecuritygroup
- --description – A description for this security group.
  Example: "My new security group"

For Linux, macOS, or Unix:

```bash
aws elasticache create-cache-security-group \
   --cache-security-group-name mysecuritygroup \
   --description "My new security group"
```

For Windows:

```bash
aws elasticache create-cache-security-group ^
   --cache-security-group-name mysecuritygroup ^
   --description "My new security group"
```

For more information, see create-cache-security-group.

Creating a Security Group (ElastiCache API)

Using the ElastiCache API operation CreateCacheSecurityGroup with the following parameters:

- CacheSecurityGroupName – The name of the security group you are creating.
  Example: mysecuritygroup
- Description – A URL encoded description for this security group.
  Example: My%20security%20group
Example

Line breaks are added for ease of reading.

https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateCacheSecurityGroup
  &CacheSecurityGroupName=mysecuritygroup
  &Description=My%20security%20group
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T220302Z
  &X-Amz-Algorithm=AWS4-HMAC-SHA256
  &X-Amz-Date=20150202T220302Z
  &X-Amz-SignedHeaders=Host
  &X-Amz-Expires=20150202T220302Z
  &X-Amz-Credential=<credential>
  &X-Amz-Signature=<signature>
Listing Available Security Groups

This topic is relevant to you only if you are not running in an Amazon VPC. If you are running in an Amazon VPC, see Amazon VPCs and ElastiCache Security (p. 198).

You can list which security groups have been created for your AWS account.

The following procedures show you how to list the available security groups for your AWS account.

**Listing Available Security Groups (Console)**

2. In the navigation pane, choose **Security Groups**.
   
   The available security groups appear in the **Security Groups** list.

**Listing Available Security Groups (AWS CLI)**

At a command prompt, use the `describe-cache-security-groups` command to list all available security groups for your AWS account.

```bash
aws elasticache describe-cache-security-groups
```

JSON output from this command will look something like this.

```json
{
  "Marker": "Marker",
  "CacheSecurityGroups": [
    {
      "OwnerId": "OwnerId",
      "CacheSecurityGroupName": "CacheSecurityGroupName",
      "Description": "Description",
      "EC2SecurityGroups": [
        {
          "Status": "Status",
          "EC2SecurityGroupName": "EC2SecurityGroupName",
          "EC2SecurityGroupOwnerId": "EC2SecurityGroupOwnerId"
        }
      ]
    }
  ]
}
```

For more information, see `describe-cache-security-groups`.

**Listing Available Security Groups (ElastiCache API)**

Using the ElastiCache API, call `DescribeCacheSecurityGroups`.

**Example**

Line breaks are added for ease of reading.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheSecurityGroups
&MaxRecords=100
&Version=2015-02-02
```

API Version 2015-02-02

225
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SignatureVersion</td>
<td>4</td>
</tr>
<tr>
<td>SignatureMethod</td>
<td>HmacSHA256</td>
</tr>
<tr>
<td>Timestamp</td>
<td>20150202T220302Z</td>
</tr>
<tr>
<td>X-Amz-Algorithm</td>
<td>AWS4-HMAC-SHA256</td>
</tr>
<tr>
<td>X-Amz-Date</td>
<td>20150202T220302Z</td>
</tr>
<tr>
<td>X-Amz-SignedHeaders</td>
<td>Host</td>
</tr>
<tr>
<td>X-Amz-Expires</td>
<td>20150202T220302Z</td>
</tr>
<tr>
<td>X-Amz-Credential</td>
<td>&lt;credential&gt;</td>
</tr>
<tr>
<td>X-Amz-Signature</td>
<td>&lt;signature&gt;</td>
</tr>
</tbody>
</table>
Viewing a Security Group

This topic is relevant to you only if you are not running in an Amazon VPC. If you are running in an Amazon VPC, see Amazon VPCs and ElastiCache Security (p. 198).

You can view detailed information about your security group.

The following procedures show you how to view the properties of a security group using the ElastiCache console, AWS CLI, and ElastiCache API.

Viewing a Security Group (Console)

2. In the navigation pane, choose Security Groups.

   The available cache security groups appear in the Security Groups list.
3. Choose a cache security group from the Security Groups list.

   The list of authorizations defined for the security group appears in the detail section at the bottom of the window.

Viewing a Security Group (AWS CLI)

At the command prompt, use the AWS CLI describe-cache-security-groups command with the name of the security group you want to view.

• --cache-security-group-name – the name of the security group to return details for.

```
aws elasticache describe-cache-security-groups --cache-security-group-name mysecuritygroup
```

JSON output from this command will look something like this.

```
{
   "CacheSecurityGroup": {
      "OwnerId": "OwnerId",
      "CacheSecurityGroupName": "CacheSecurityGroupName",
      "Description": "Description",
      "EC2SecurityGroups": [
         {
            "Status": "Status",
            "EC2SecurityGroupName": "EC2SecurityGroupName",
            "EC2SecurityGroupOwnerId": "EC2SecurityGroupOwnerId"
         }
      ]
   }
}
```

For more information, see describe-cache-security-groups.

Viewing a Security Group (ElastiCache API)

Using the ElastiCache API, call DescribeCacheSecurityGroups with the name of the security group you want to view.

• CacheSecurityGroupName – the name of the cache security group to return details for.
Example

Line breaks are added for ease of reading.

https://elasticache.amazonaws.com/
   ?Action=DescribeCacheSecurityGroups
   &CacheSecurityGroupName=mysecuritygroup
   &Version=2015-02-02
   &SignatureVersion=4
   &SignatureMethod=HmacSHA256
   &Timestamp=20150202T220302Z
   &X-Amz-Algorithm=AWS4-HMAC-SHA256
   &X-Amz-Date=20150202T220302Z
   &X-Amz-SignedHeaders=Host
   &X-Amz-Expires=20150202T220302Z
   &X-Amz-Credential=<credential>
   &X-Amz-Signature=<signature>
Authorizing Network Access to an Amazon EC2 Security Group

This topic is relevant to you only if you are not running in an Amazon VPC. If you are running in an Amazon VPC, see Amazon VPCs and ElastiCache Security (p. 198).

If you want to access your cluster from an Amazon EC2 instance, you must grant access to the Amazon EC2 security group that the EC2 instance belongs to. The following procedures show you how to grant access to an Amazon EC2 Security Group.

Important

- Authorizing an Amazon EC2 security group only grants access to your clusters from all EC2 instances belonging to the Amazon EC2 security group.
- It takes approximately one minute for changes to access permissions to take effect.

Authorizing Network Access to an Amazon EC2 Security Group (Console)

2. In the navigation pane, choose Security Groups.
3. In the Security Groups list, choose the box to the left of the security group that you want to grant access to.
4. At the bottom of the window, in the EC2 Security Group Name list, choose your Amazon EC2 security group.
5. Choose Add.

Authorizing Network Access to an Amazon EC2 Security Group (AWS CLI)

At a command prompt, use the authorize-cache-security-group-ingress command to grant access to an Amazon EC2 security group with the following parameters.

- --cache-security-group-name – the name of the security group you are granting Amazon EC2 access to.
- --ec2-security-group-name – the name of the Amazon EC2 security group that the Amazon EC2 instance belongs to.
- --ec2-security-group-owner-id – the id of the owner of the Amazon EC2 security group.

Example

For Linux, macOS, or Unix:

```bash
aws elasticache authorize-cache-security-group-ingress \
   --cache-security-group-name default \
   --ec2-security-group-name myec2group \
   --ec2-security-group-owner-id 987654321021
```

For Windows:

```bash
aws elasticache authorize-cache-security-group-ingress ^
   --cache-security-group-name default ^
   --ec2-security-group-name myec2group ^
   --ec2-security-group-owner-id 987654321021
```
The command should produce output similar to the following:

```json
{
  "CacheSecurityGroup": {
    "OwnerId": "OwnerId",
    "CacheSecurityGroupName": "CacheSecurityGroupName",
    "Description": "Description",
    "EC2SecurityGroups": [
      {
        "Status": "available",
        "EC2SecurityGroupName": "EC2SecurityGroupName",
        "EC2SecurityGroupOwnerId": "EC2SecurityGroupOwnerId"
      }
    ]
  }
}
```

For more information, see `authorize-cache-security-group-ingress`.

**Authorizing Network Access to an Amazon EC2 Security Group (ElastiCache API)**

Using the ElastiCache API, call `AuthorizeCacheSecurityGroupIngress` with the following parameters:

- **CacheSecurityGroupName** – the name of the security group you are granting Amazon EC2 access to.
- **EC2SecurityGroupName** – the name of the Amazon EC2 security group that the Amazon EC2 instance belongs to.
- **EC2SecurityGroupOwnerId** – the id of the owner of the Amazon EC2 security group.

**Example**

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=AuthorizeCacheSecurityGroupIngress
  &EC2SecurityGroupOwnerId=987654321021
  &EC2SecurityGroupName=myec2group
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T220302Z
  &X-Amz-Algorithm=AWS4-HMAC-SHA256
  &X-Amz-Date=20150202T220302Z
  &X-Amz-SignedHeaders=Host
  &X-Amz-Expires=20150202T220302Z
  &X-Amz-Credential=<credential>
  &X-Amz-Signature=<signature>
```

For more information, see `AuthorizeCacheSecurityGroupIngress`.

**Identity and Access Management in Amazon ElastiCache**

Access to Amazon ElastiCache requires credentials that AWS can use to authenticate your requests. Those credentials must have permissions to access AWS resources, such as an ElastiCache cache cluster or an Amazon Elastic Compute Cloud (Amazon EC2) instance. The following sections provide details on how
you can use AWS Identity and Access Management (IAM) and ElastiCache to help secure your resources by controlling who can access them.

- Authentication (p. 231)
- Access Control (p. 232)

**Authentication**

You can access AWS as any of the following types of identities:

- **AWS account root user** – When you first create an AWS account, you begin with a single sign-in identity that has complete access to all AWS services and resources in the account. This identity is called the AWS account root user and is accessed by signing in with the email address and password that you used to create the account. We strongly recommend that you do not use the root user for your everyday tasks, even the administrative ones. Instead, adhere to the best practice of using the root user only to create your first IAM user. Then securely lock away the root user credentials and use them to perform only a few account and service management tasks.

- **IAM user** – An IAM user is an identity within your AWS account that has specific custom permissions (for example, permissions to create a cluster in ElastiCache). You can use an IAM user name and password to sign in to secure AWS webpages like the AWS Management Console, AWS Discussion Forums, or the AWS Support Center.

In addition to a user name and password, you can also generate access keys for each user. You can use these keys when you access AWS services programmatically, either through one of the several SDKs or by using the AWS Command Line Interface (CLI). The SDK and CLI tools use the access keys to cryptographically sign your request. If you don’t use AWS tools, you must sign the request yourself. ElastiCache supports Signature Version 4, a protocol for authenticating inbound API requests. For more information about authenticating requests, see Signature Version 4 Signing Process in the AWS General Reference.

- **IAM role** – An IAM role is an IAM identity that you can create in your account that has specific permissions. An IAM role is similar to an IAM user in that it is an AWS identity with permissions policies that determine what the identity can and cannot do in AWS. However, instead of being uniquely associated with one person, a role is intended to be assumable by anyone who needs it. Also, a role does not have standard long-term credentials such as a password or access keys associated with it. Instead, when you assume a role, it provides you with temporary security credentials for your role session. IAM roles with temporary credentials are useful in the following situations:

- **Federated user access** – Instead of creating an IAM user, you can use existing identities from AWS Directory Service, your enterprise user directory, or a web identity provider. These are known as federated users. AWS assigns a role to a federated user when access is requested through an identity provider. For more information about federated users, see Federated Users and Roles in the IAM User Guide.

- **AWS service access** – A service role is an IAM role that a service assumes to perform actions in your account on your behalf. When you set up some AWS service environments, you must define a role for the service to assume. This service role must include all the permissions that are required for the service to access the AWS resources that it needs. Service roles vary from service to service, but many allow you to choose your permissions as long as you meet the documented requirements for that service. Service roles provide access only within your account and cannot be used to grant access to services in other accounts. You can create, modify, and delete a service role from within...
IAM. For example, you can create a role that allows Amazon Redshift to access an Amazon S3 bucket on your behalf and then load data from that bucket into an Amazon Redshift cluster. For more information, see Creating a Role to Delegate Permissions to an AWS Service in the IAM User Guide.

- **Applications running on Amazon EC2** – You can use an IAM role to manage temporary credentials for applications that are running on an EC2 instance and making AWS CLI or AWS API requests. This is preferable to storing access keys within the EC2 instance. To assign an AWS role to an EC2 instance and make it available to all of its applications, you create an instance profile that is attached to the instance. An instance profile contains the role and enables programs that are running on the EC2 instance to get temporary credentials. For more information, see Using an IAM Role to Grant Permissions to Applications Running on Amazon EC2 Instances in the IAM User Guide.

**Access Control**

You can have valid credentials to authenticate your requests, but unless you have permissions you cannot create or access Amazon ElastiCache resources. For example, you must have permissions to create an ElastiCache cache cluster.

The following sections describe how to manage permissions for Amazon ElastiCache. We recommend that you read the overview first.

- **Overview of Managing Access Permissions to Your ElastiCache Resources** (p. 233)
- **Using Identity-Based Policies (IAM Policies) for Amazon ElastiCache** (p. 237)
Overview of Managing Access Permissions to Your ElastiCache Resources

Every AWS resource is owned by an AWS account, and permissions to create or access a resource are governed by permissions policies. An account administrator can attach permissions policies to IAM identities (that is, users, groups, and roles). In addition, some services (such as AWS Lambda) also support attaching permissions policies to resources.

**Note**

An account administrator (or administrator user) is a user with administrator privileges. For more information, see IAM Best Practices in the IAM User Guide.

When granting permissions, you decide who is getting the permissions. You also decide the resources they get permissions for and the specific actions that you want to allow on those resources.

**Topics**

- Amazon ElastiCache Resources and Operations (p. 233)
- Understanding Resource Ownership (p. 233)
- Managing Access to Resources (p. 234)
- Specifying Policy Elements: Actions, Effects, Resources, and Principals (p. 235)
- Specifying Conditions in a Policy (p. 235)

Amazon ElastiCache Resources and Operations

In Amazon ElastiCache, the primary resource is a cache cluster.

These resources have unique Amazon Resource Names (ARNs) associated with them as shown in the following table.

<table>
<thead>
<tr>
<th>Resource Type</th>
<th>ARN Format</th>
</tr>
</thead>
</table>

ElastiCache provides a set of operations to work with ElastiCache resources. For a list of available operations, see Amazon ElastiCache Actions.

Understanding Resource Ownership

A resource owner is the AWS account that created the resource. That is, the resource owner is the AWS account of the principal entity that authenticates the request that creates the resource. A principal entity can be the root account, an IAM user, or an IAM role). The following examples illustrate how this works:

- Suppose that you use the root account credentials of your AWS account to create a cache cluster. In this case, your AWS account is the owner of the resource. In ElastiCache, the resource is the cache cluster.
- Suppose that you create an IAM user in your AWS account and grant permissions to create a cache cluster to that user. In this case, the user can create a cache cluster. However, your AWS account, to which the user belongs, owns the cache cluster resource.
- Suppose that you create an IAM role in your AWS account with permissions to create a cache cluster. In this case, anyone who can assume the role can create a cache cluster. Your AWS account, to which the role belongs, owns the cache cluster resource.
Managing Access to Resources

A permissions policy describes who has access to what. The following section explains the available options for creating permissions policies.

Note
This section discusses using IAM in the context of Amazon ElastiCache. It doesn't provide detailed information about the IAM service. For complete IAM documentation, see What Is IAM? in the IAM User Guide. For information about IAM policy syntax and descriptions, see AWS IAM Policy Reference in the IAM User Guide.

Policies attached to an IAM identity are referred to as identity-based policies (IAM policies). Policies attached to a resource are referred to as resource-based policies. Amazon ElastiCache supports only identity-based policies (IAM policies).

Topics
- Identity-Based Policies (IAM Policies) (p. 234)
- Resource-Based Policies (p. 235)

Identity-Based Policies (IAM Policies)

You can attach policies to IAM identities. For example, you can do the following:

• Attach a permissions policy to a user or a group in your account – An account administrator can use a permissions policy that is associated with a particular user to grant permissions. In this case, the permissions are for that user to create an ElastiCache resource, such as a cache cluster, parameter group, or security group.

• Attach a permissions policy to a role (grant cross-account permissions) – You can attach an identity-based permissions policy to an IAM role to grant cross-account permissions. For example, the administrator in Account A can create a role to grant cross-account permissions to another AWS account (for example, Account B) or an AWS service as follows:
  1. Account A administrator creates an IAM role and attaches a permissions policy to the role that grants permissions on resources in Account A.
  2. Account A administrator attaches a trust policy to the role identifying Account B as the principal who can assume the role.
  3. Account B administrator can then delegate permissions to assume the role to any users in Account B. Doing this allows users in Account B to create or access resources in Account A. In some cases, you might want to grant an AWS service permissions to assume the role. To support this approach, the principal in the trust policy can also be an AWS service principal.

For more information about using IAM to delegate permissions, see Access Management in the IAM User Guide.

The following is an example policy that allows a user to perform the DescribeCacheClusters action for your AWS account. In the current implementation, ElastiCache doesn't support identifying specific resources using the resource ARNs for API actions. (This approach is also referred to as resource-level permissions). Thus, you must specify a wildcard character (*).

```json
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Sid": "DescribeCacheClusters",
         "Effect": "Allow",
         "Action": ["elasticache:DescribeCacheClusters"]
      }
   ]
}
```
Overview of Managing Access

For more information about using identity-based policies with ElastiCache, see Using Identity-Based Policies (IAM Policies) for Amazon ElastiCache (p. 237). For more information about users, groups, roles, and permissions, see Identities (Users, Groups, and Roles in the IAM User Guide).

Resource-Based Policies

Other services, such as Amazon S3, also support resource-based permissions policies. For example, you can attach a policy to an S3 bucket to manage access permissions to that bucket. Amazon ElastiCache doesn't support resource-based policies.

Specifying Policy Elements: Actions, Effects, Resources, and Principals

For each Amazon ElastiCache resource (see Amazon ElastiCache Resources and Operations (p. 233)), the service defines a set of API operations (see Actions). To grant permissions for these API operations, ElastiCache defines a set of actions that you can specify in a policy. For example, for the ElastiCache snapshot resource, the following actions are defined: CreateCacheCluster, DeleteCacheCluster, and DescribeCacheCluster. Performing an API operation can require permissions for more than one action.

The following are the most basic policy elements:

- **Resource** – In a policy, you use an Amazon Resource Name (ARN) to identify the resource to which the policy applies. For ElastiCache resources, you always use the wildcard character (*) in IAM policies. For more information, see Amazon ElastiCache Resources and Operations (p. 233).

- **Action** – You use action keywords to identify resource operations that you want to allow or deny. For example, depending on the specified Effect, the elasticache:CreateCacheCluster permission allows or denies the user permissions to perform the Amazon ElastiCache CreateCacheCluster operation.

- **Effect** – You specify the effect when the user requests the specific action—this can be either allow or deny. If you don't explicitly grant access to (allow) a resource, access is implicitly denied. You can also explicitly deny access to a resource. For example, you might do this to make sure that a user can't access a resource, even if a different policy grants access.

- **Principal** – In identity-based policies (IAM policies), the user that the policy is attached to is the implicit principal. For resource-based policies, you specify the user, account, service, or other entity that you want to receive permissions (applies to resource-based policies only). ElastiCache doesn't support resource-based policies.

To learn more about IAM policy syntax and descriptions, see AWS IAM Policy Reference in the IAM User Guide.

For a table showing all of the Amazon ElastiCache API actions, see ElastiCache API Permissions: Actions, Resources, and Conditions Reference (p. 248).

Specifying Conditions in a Policy

When you grant permissions, you can use the IAM policy language to specify the conditions when a policy should take effect. For example, you might want a policy to be applied only after a specific date. For more information about specifying conditions in a policy language, see Condition in the IAM User Guide.
To express conditions, you use predefined condition keys. There are no condition keys specific to Amazon ElastiCache. However, there are AWS-wide condition keys that you can use as appropriate. For a complete list of AWS-wide keys, see Available Keys for Conditions in the IAM User Guide.
Using Identity-Based Policies (IAM Policies) for Amazon ElastiCache

This topic provides examples of identity-based policies in which an account administrator can attach permissions policies to IAM identities (that is, users, groups, and roles).

Important
We recommend that you first read the topics that explain the basic concepts and options to manage access to Amazon ElastiCache resources. For more information, see Overview of Managing Access Permissions to Your ElastiCache Resources (p. 233).

The sections in this topic cover the following:

- Permissions Required to Use the Amazon ElastiCache Console (p. 238)
- AWS-Managed (Predefined) Policies for Amazon ElastiCache (p. 238)
- Customer-Managed Policy Examples (p. 239)

The following shows an example of a permissions policy.

```json
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "AllowClusterPermissions",
        "Effect": "Allow",
        "Action": [
            "elasticache:CreateCacheCluster",
            "elasticache:CreateReplicationGroup",
            "elasticache:DescribeCacheClusters",
            "elasticache:ModifyCacheCluster",
            "elasticache:RebootCacheCluster"
        ],
        "Resource": "*"
    }]
}
```

The policy has two statements:

- The first statement grants permissions for the Amazon ElastiCache actions (elasticache:CreateCacheCluster, elasticache:DescribeCacheClusters, elasticache:ModifyCacheCluster, and elasticache:RebootCacheCluster) on any cache cluster owned by the account. Currently, Amazon ElastiCache doesn't support permissions for actions at the resource-level. Therefore, the policy specifies a wildcard character (*) as the Resource value.

- The second statement grants permissions for the IAM action (iam:PassRole) on IAM roles. The wildcard character (*) at the end of the Resource value means that the statement allows permission for the iam:PassRole action on any IAM role. To limit this permission to a specific role, replace the wildcard character (*) in the resource ARN with the specific role name.

The policy doesn't specify the Principal element because in an identity-based policy you don't specify the principal who gets the permission. When you attach policy to a user, the user is the implicit principal. When you attach a permissions policy to an IAM role, the principal identified in the role's trust policy gets the permissions.

For a table showing all of the Amazon ElastiCache API actions and the resources that they apply to, see ElastiCache API Permissions: Actions, Resources, and Conditions Reference (p. 248).
Permissions Required to Use the Amazon ElastiCache Console

The permissions reference table lists the Amazon ElastiCache API operations and shows the required permissions for each operation. For more information about ElastiCache API operations, see ElastiCache API Permissions: Actions, Resources, and Conditions Reference (p. 248).

To use the Amazon ElastiCache console, first grant permissions for additional actions as shown in the following permissions policy.

```json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "MinPermsForECConsole",
            "Effect": "Allow",
            "Action": [
                "elasticache:Describe*",
                "elasticache:List*",
                "ec2:DescribeAvailabilityZones",
                "ec2:DescribeVpcs",
                "ec2:DescribeAccountAttributes",
                "ec2:DescribeSecurityGroups",
                "cloudwatch:GetMetricStatistics",
                "cloudwatch:DescribeAlarms",
                "s3:ListAllMyBuckets",
                "sns:ListTopics",
                "sns:ListSubscriptions"
            ],
            "Resource": "*"
        }
    ]
}
```

The ElastiCache console needs these additional permissions for the following reasons:

- Permissions for the ElastiCache actions enable the console to display ElastiCache resources in the account.
- The console needs permissions for the ec2 actions to query Amazon EC2 so it can display Availability Zones, VPCs, security groups, and account attributes.
- The permissions for cloudwatch actions enable the console to retrieve Amazon CloudWatch metrics and alarms, and display them in the console.
- The permissions for sns actions enable the console to retrieve Amazon Simple Notification Service (Amazon SNS) topics and subscriptions, and display them in the console.

AWS-Managed (Predefined) Policies for Amazon ElastiCache

AWS addresses many common use cases by providing standalone IAM policies that are created and administered by AWS. Managed policies grant necessary permissions for common use cases so you can avoid having to investigate what permissions are needed. For more information, see AWS Managed Policies in the IAM User Guide.

The following AWS managed policies, which you can attach to users in your account, are specific to ElastiCache:

- **AmazonElastiCacheReadOnlyAccess** - Grants read-only access to Amazon ElastiCache resources.
- **AmazonElastiCacheFullAccess** - Grants full access to Amazon ElastiCache resources.
Note
You can review these permissions policies by signing in to the IAM console and searching for specific policies there.

You can also create your own custom IAM policies to allow permissions for Amazon ElastiCache API actions. You can attach these custom policies to the IAM users or groups that require those permissions.

Customer-Managed Policy Examples

If you are not using a default policy and choose to use a custom-managed policy, ensure one of two things. Either you should have permissions to call `iam:createServiceLinkedRole` (for more information, see Example 5: Allow a User to Call IAM CreateServiceLinkedRole API (p. 241)). Or you should have created an ElastiCache service-linked role.

When combined with the minimum permissions needed to use the Amazon ElastiCache console, the example policies in this section grant additional permissions. The examples are also relevant to the AWS SDKs and the AWS CLI. For more information about what permissions are needed to use the ElastiCache console, see Permissions Required to Use the Amazon ElastiCache Console (p. 238).

For instructions on setting up IAM users and groups, see Creating Your First IAM User and Administrators Group in the IAM User Guide.

Important
Always test your IAM policies thoroughly before using them in production. Some ElastiCache actions that appear simple can require other actions to support them when you are using the ElastiCache console. For example, `elasticache:CreateCacheCluster` grants permissions to create ElastiCache cache clusters. However, to perform this operation, the ElastiCache console uses a number of `Describe` and `List` actions to populate console lists.

Examples

- Example 1: Allow a User to Create and Manage Security Groups (p. 239)
- Example 2: Allow a User Read-Only Access to ElastiCache Resources (p. 240)
- Example 3: Allow a User to Perform Common ElastiCache System Administrator Tasks (p. 240)
- Example 4: Allow a User to Access All ElastiCache API Actions (p. 240)
- Example 5: Allow a User to Call IAM CreateServiceLinkedRole API (p. 241)

Example 1: Allow a User to Create and Manage Security Groups

The following policy grants permissions for the security group's specific ElastiCache actions. Typically, you attach this type of permissions policy to the system administrators group.

```
{
    "Version": "2012-10-17",
    "Statement":[
        {
            "Sid": "SecGrpAllows",
            "Effect": "Allow",
            "Action": [
                "elasticache:CreateCacheSecurityGroup",
                "elasticache:DeleteCacheSecurityGroup",
                "elasticache:DescribeCacheSecurityGroup",
                "elasticache:AuthorizeCacheSecurityGroupIngress",
                "elasticache:RevokeCacheSecurityGroupIngress"
            ],
            "Resource": "*"
        }
    ]
}
```
Example 2: Allow a User Read-Only Access to ElastiCache Resources

The following policy grants permissions ElastiCache actions that allow a user to list resources. Typically, you attach this type of permissions policy to a managers group.

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "ECUnrestricted",
        "Effect": "Allow",
        "Action": [ "elasticache:Describe*", "elasticache:List*" ],
        "Resource": "*"
    }]
}
```

Example 3: Allow a User to Perform Common ElastiCache System Administrator Tasks

Common system administrator tasks include modifying cache clusters, parameters, and parameter groups. A system administrator may also want to get information about the ElastiCache events. The following policy grants a user permissions to perform ElastiCache actions for these common system administrator tasks. Typically, you attach this type of permissions policy to the system administrators group.

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "ECAllowSpecific",
        "Effect": "Allow",
        "Resource": "*"
    }]
}
```

Example 4: Allow a User to Access All ElastiCache API Actions

The following policy allows a user to access all ElastiCache actions. We recommend that you grant this type of permissions policy only to an administrator user.

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "ECAllowSpecific",
        "Effect": "Allow",
        "Action": [ "elasticache:*" ],
```
Example 5: Allow a User to Call IAM CreateServiceLinkedRole API

The following policy allows user to call the IAM CreateServiceLinkedRole API. We recommend that you grant this type of permissions policy to the user who invokes mutative ElastiCache operations.

```json
{
"Version":"2012-10-17",
"Statement":[
{
   "Sid":"CreateSLRAllows",
   "Effect":"Allow",
   "Action":[
   "iam:CreateServiceLinkedRole"
   ],
   "Resource":"*",
   "Condition":{
   "StringLike":{
   "iam:AWSServiceName":"elasticache.amazonaws.com"
   }
   }
   }
]
}
```

Using Service-Linked Roles for Amazon ElastiCache

Amazon ElastiCache uses AWS Identity and Access Management (IAM) service-linked roles. A service-linked role is a unique type of IAM role that is linked directly to an AWS service, such as Amazon ElastiCache. Amazon ElastiCache service-linked roles are predefined by Amazon ElastiCache. They include all the permissions that the service requires to call AWS services on behalf of your clusters.

A service-linked role makes setting up Amazon ElastiCache easier because you don't have to manually add the necessary permissions. The roles already exist within your AWS account but are linked to Amazon ElastiCache use cases and have predefined permissions. Only Amazon ElastiCache can assume these roles, and only these roles can use the predefined permissions policy. You can delete the roles only after first deleting their related resources. This protects your Amazon ElastiCache resources because you can’t inadvertently remove necessary permissions to access the resources.

For information about other services that support service-linked roles, see AWS Services That Work with IAM and look for the services that have Yes in the Service-Linked Role column. Choose a Yes with a link to view the service-linked role documentation for that service.

Contents

- Service-Linked Role Permissions for Amazon ElastiCache (p. 242)
- Creating a Service-Linked Role (IAM) (p. 243)
  - Creating a Service-Linked Role (IAM Console) (p. 243)
  - Creating a Service-Linked Role (IAM CLI) (p. 243)
  - Creating a Service-Linked Role (IAM API) (p. 243)
- Editing the Description of a Service-Linked Role for Amazon ElastiCache (p. 244)
  - Editing a Service-Linked Role Description (IAM Console) (p. 244)
  - Editing a Service-Linked Role Description (IAM CLI) (p. 244)
Amazon ElastiCache uses the service-linked role named **AWSServiceRoleForElastiCache** — This policy allows ElastiCache to manage AWS resources on your behalf as necessary for managing your cache.

The AWSForElastiCache service-linked role permissions policy allows Amazon ElastiCache to complete the following actions on the specified resources:

```
Permission policy:
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:AuthorizeSecurityGroupIngress",
        "ec2:CreateNetworkInterface",
        "ec2:CreateSecurityGroup",
        "ec2:DeleteNetworkInterface",
        "ec2:DeleteSecurityGroup",
        "ec2:DescribeAvailabilityZones",
        "ec2:DescribeNetworkInterfaces",
        "ec2:DescribeSecurityGroups",
        "ec2:DescribeSubnets",
        "ec2:DescribeVpcs",
        "ec2:ModifyNetworkInterfaceAttribute",
        "ec2:RevokeSecurityGroupIngress",
      ],
      "Resource": "*"
    }
  ]
}
```

**To allow an IAM entity to create AWSForElastiCache service-linked roles**

Add the following policy statement to the permissions for that IAM entity:

```
{
  "Effect": "Allow",
  "Action": [
    "iam:CreateServiceLinkedRole",
    "iam:PutRolePolicy"
  ],
  "Resource": "arn:aws:iam::*:role/aws-service-role/elasticache.amazonaws.com/AWSForElastiCache",
  "Condition": {"StringLike": {"iam:AWSManagedPolicyName": "elasticache.amazonaws.com"}}
}
```

**To allow an IAM entity to delete AWSForElastiCache service-linked roles**

Add the following policy statement to the permissions for that IAM entity:

```
```
Alternatively, you can use an AWS managed policy to provide full access to Amazon ElastiCache.

Creating a Service-Linked Role (IAM)

You can create a service-linked role using the IAM console, CLI, or API.

Creating a Service-Linked Role (IAM Console)

You can use the IAM console to create a service-linked role.

To create a service-linked role (console)

1. Sign in to the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane of the IAM console, choose **Roles**. Then choose **Create new role**.
3. Expand the **AWS service-linked role** section, and then select the service that you want to allow to assume this new service-linked role.
4. Next to the **AWSServiceRoleForElastiCache** service-linked role, choose **Select**.
5. For **Role name**, type a suffix to add to the service-linked role default name. This suffix helps you identify the purpose of this role. Role names must be unique within your AWS account. They are not distinguished by case. For example, you cannot create roles named both `<service-linked-role-name>_SAMPLE` and `<service-linked-role-name>_sample`. Because various entities might reference the role, you cannot edit the name of the role after it has been created.
6. (Optional) For **Role description**, edit the description for the new service-linked role.
7. Review the role and then choose **Create role**.

Creating a Service-Linked Role (IAM CLI)

You can use IAM operations from the AWS Command Line Interface to create a service-linked role. This role can include the trust policy and inline policies that the service needs to assume the role.

To create a service-linked role (CLI)

Use the following operation:

```
$ aws iam create-service-linked-role --aws-service-name elasticache.amazonaws.com
```

Creating a Service-Linked Role (IAM API)

You can use the IAM API to create a service-linked role. This role can contain the trust policy and inline policies that the service needs to assume the role.

To create a service-linked role (API)

```json
{
   "Effect": "Allow",
   "Action": [
      "iam:DeleteServiceLinkedRole",
      "iam:GetServiceLinkedRoleDeletionStatus"
   ],
   "Resource": "arn:aws:iam::*:role/aws-service-role/elasticache.amazonaws.com/AWSServiceRoleForElastiCache*",
   "Condition": {
      "StringLike": {"iam:AWSServiceName": "elasticache.amazonaws.com"}
   }
}
```
Use the CreateServiceLinkedRole API call. In the request, specify a service name of elasticache.amazonaws.com.

## Editing the Description of a Service-Linked Role for Amazon ElastiCache

Amazon ElastiCache does not allow you to edit the AWSServiceRoleForElastiCache service-linked role. After you create a service-linked role, you cannot change the name of the role because various entities might reference the role. However, you can edit the description of the role using IAM.

### Editing a Service-Linked Role Description (IAM Console)

You can use the IAM console to edit a service-linked role description.

**To edit the description of a service-linked role (console)**

1. In the navigation pane of the IAM console, choose **Roles**.
2. Choose the name of the role to modify.
3. To the far right of **Role description**, choose **Edit**.
4. Enter a new description in the box and choose **Save**.

### Editing a Service-Linked Role Description (IAM CLI)

You can use IAM operations from the AWS Command Line Interface to edit a service-linked role description.

**To change the description of a service-linked role (CLI)**

1. (Optional) To view the current description for a role, use the AWS CLI for IAM operation `get-role`.
   
   **Example**
   
   ```bash
   $ aws iam get-role --role-name AWSServiceRoleForElastiCache
   ```

   Use the role name, not the ARN, to refer to roles with the CLI operations. For example, if a role has the following ARN: `arn:aws:iam::123456789012:role/myrole`, refer to the role as `myrole`.

2. To update a service-linked role’s description, use the AWS CLI for IAM operation `update-role-description`.

   For Linux, macOS, or Unix:
   
   ```bash
   $ aws iam update-role-description \
       --role-name AWSServiceRoleForElastiCache \
       --description "new description"
   ```

   For Windows:
   
   ```bash
   $ aws iam update-role-description ^
       --role-name AWSServiceRoleForElastiCache ^
       --description "new description"
   ```

### Editing a Service-Linked Role Description (IAM API)

You can use the IAM API to edit a service-linked role description.

API Version 2015-02-02
To change the description of a service-linked role (API)

1. (Optional) To view the current description for a role, use the IAM API operation GetRole.

   Example

   ```
   https://iam.amazonaws.com/
   ?Action=GetRole
   &RoleName=AWSServiceRoleForElastiCache
   &Version=2010-05-08
   &AUTHPARAMS
   ```

2. To update a role’s description, use the IAM API operation UpdateRoleDescription.

   Example

   ```
   https://iam.amazonaws.com/
   ?Action=UpdateRoleDescription
   &RoleName=AWSServiceRoleForElastiCache
   &Version=2010-05-08
   &Description="New description"
   ```

Deleting a Service-Linked Role for Amazon ElastiCache

If you no longer need to use a feature or service that requires a service-linked role, we recommend that you delete that role. That way you don’t have an unused entity that is not actively monitored or maintained. However, you must clean up your service-linked role before you can delete it.

Amazon ElastiCache does not delete the service-linked role for you.

Cleaning Up a Service-Linked Role

Before you can use IAM to delete a service-linked role, first confirm that the role has no resources—clusters—associated with it.

To check whether the service-linked role has an active session in the IAM console

1. Sign in to the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane of the IAM console, choose Roles. Then choose the name (not the check box) of the AWSServiceRoleForElastiCache role.
3. On the Summary page for the selected role, choose the Access Advisor tab.
4. On the Access Advisor tab, review recent activity for the service-linked role.

To delete Amazon ElastiCache resources that require AWSServiceRoleForElastiCache (console)

- To delete a cluster, see the following:
  - Using the AWS Management Console (p. 105)
  - Using the AWS CLI (p. 105)
  - Using the ElastiCache API (p. 106)

Deleting a Service-Linked Role (IAM Console)

You can use the IAM console to delete a service-linked role.
To delete a service-linked role (console)

1. Sign in to the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane of the IAM console, choose Roles. Then select the check box next to the role name that you want to delete, not the name or row itself.
3. For Role actions at the top of the page, choose Delete role.
4. In the confirmation dialog box, review the service last accessed data, which shows when each of the selected roles last accessed an AWS service. This helps you to confirm whether the role is currently active. If you want to proceed, choose Yes, Delete to submit the service-linked role for deletion.
5. Watch the IAM console notifications to monitor the progress of the service-linked role deletion. Because the IAM service-linked role deletion is asynchronous, after you submit the role for deletion, the deletion task can succeed or fail. If the task fails, you can choose View details or View Resources from the notifications to learn why the deletion failed.

Deleting a Service-Linked Role (IAM CLI)

You can use IAM operations from the AWS Command Line Interface to delete a service-linked role.

To delete a service-linked role (CLI)

1. If you don't know the name of the service-linked role that you want to delete, enter the following command. This command lists the roles and their Amazon Resource Names (ARNs) in your account.

   ```bash
   $ aws iam get-role --role-name role-name
   ```

   Use the role name, not the ARN, to refer to roles with the CLI operations. For example, if a role has the ARN arn:aws:iam::123456789012:role/myrole, you refer to the role as myrole.

2. Because a service-linked role cannot be deleted if it is being used or has associated resources, you must submit a deletion request. That request can be denied if these conditions are not met. You must capture the DeletionTaskId from the response to check the status of the deletion task. Enter the following to submit a service-linked role deletion request.

   ```bash
   $ aws iam delete-service-linked-role --role-name role-name
   ```

3. Enter the following to check the status of the deletion task.

   ```bash
   $ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
   ```

   The status of the deletion task can be NOT_STARTED, IN_PROGRESS, SUCCEEDED, or FAILED. If the deletion fails, the call returns the reason that it failed so that you can troubleshoot.

Deleting a Service-Linked Role (IAM API)

You can use the IAM API to delete a service-linked role.

To delete a service-linked role (API)

1. To submit a deletion request for a service-linked roll, call DeleteServiceLinkedRole. In the request, specify a role name.

   Because a service-linked role cannot be deleted if it is being used or has associated resources, you must submit a deletion request. That request can be denied if these conditions are not met. You must capture the DeletionTaskId from the response to check the status of the deletion task.
2. To check the status of the deletion, call `GetServiceLinkedRoleDeletionStatus`. In the request, specify the `DeletionTaskId`.

The status of the deletion task can be `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, or `FAILED`. If the deletion fails, the call returns the reason that it failed so that you can troubleshoot.
ElastiCache API Permissions: Actions, Resources, and Conditions Reference

When you set up access control (p. 232) and write permissions policies to attach to an IAM identity (identity-based policies), use the following table as a reference. The table lists each Amazon ElastiCache API operation and the corresponding actions for which you can grant permissions to perform the action. You specify the actions in the policy's Action field, and you specify a wildcard character (*) as the resource value in the policy's Resource field.

You can use AWS-wide condition keys in your ElastiCache policies to express conditions. For a complete list of AWS-wide keys, see Available Keys for Conditions in the IAM User Guide.

**Note**
To specify an action, use the elasticache: prefix followed by the API operation name (for example, elasticache:DescribeCacheClusters). For all ElastiCache actions, specify the wildcard character (*) as the resource.

### Amazon ElastiCache API and Required Permissions for Actions

#### AddTagsToResource

**Action:** elasticache:AddTagsToResource

**Resource:** *

#### AuthorizeCacheSecurityGroupIngress

**Action:** elasticache:AuthorizeCacheSecurityGroupIngress

**Resource:** *

#### CreateCacheCluster

**Actions:** elasticache:CreateCacheCluster

s3:GetObject

**Note**
If you use the SnapshotArns parameter, each member of the SnapshotArns list requires its own s3:GetObject permission with the s3 ARN as its resource.

**Resource:** *

arn:aws:s3:::my_bucket/snapshot1.rdb

Where **my_bucket/snapshot1** is an S3 bucket and snapshot that you want to create the cache cluster from.

#### CreateCacheParameterGroup

**Action:** elasticache:CreateCacheParameterGroup

**Resource:** *

#### CreateCacheSecurityGroup

**Action:** elasticache:CreateCacheSecurityGroup

**Resource:** *

#### CreateCacheSubnetGroup

**Action:** elasticache:CreateCacheSubnetGroup

**Resource:** *
Resource: *
DeleteCacheCluster
Action: elasticache:DeleteCacheCluster
Resource: *
DeleteCacheParameterGroup
Action: elasticache:DeleteCacheParameterGroup
Resource: *
DeleteCacheSecurityGroup
Action: elasticache:DeleteCacheSecurityGroup
Resource: *
DeleteCacheSubnetGroup
Action: elasticache:DeleteCacheSubnetGroup
Resource: *
DescribeCacheClusters
Action: elasticache:DescribeCacheClusters
Resource: *
DescribeCacheEngineVersions
Actions: elasticache:DescribeCacheEngineVersions
Resource: *
DescribeCacheParameterGroups
Action: elasticache:DescribeCacheParameterGroups
Resource: *
DescribeCacheParameters
Action: elasticache:DescribeCacheParameters
Resource: *
DescribeCacheSecurityGroups
Action: elasticache:DescribeCacheSecurityGroups
Resource: *
DescribeCacheSubnetGroups
Action: elasticache:DescribeCacheSubnetGroups
Resource: *
DescribeEngineDefaultParameters
Action: elasticache:DescribeEngineDefaultParameters
Resource: *
DescribeEvents
Action: elasticache:DescribeEvents
Resource: *
DescribeReservedCacheNodes
  Action: elasticache:DescribeReservedCacheNodes
Resource: *
DescribeReservedCacheNodesOfferings
  Action: elasticache:DescribeReservedCacheNodesOfferings
Resource: *
ListTagsForResource
  Action: elasticache:ListTagsForResource
Resource: *
ModifyCacheCluster
  Action: elasticache:ModifyCacheCluster
Resource: *
ModifyCacheParameterGroup
  Action: elasticache:ModifyCacheParameterGroup
Resource: *
ModifyCacheSubnetGroup
  Action: elasticache:ModifyCacheSubnetGroup
Resource: *
PurchaseReservedCacheNodesOffering
  Action: elasticache:PurchaseReservedCacheNodesOffering
Resource: *
RebootCacheCluster
  Action: elasticache:RebootCacheCluster
Resource: *
RemoveTagsFromResource
  Action: elasticache:RemoveTagsFromResource
Resource: *
ResetCacheParameterGroup
  Action: elasticache:ResetCacheParameterGroup
Resource: *
RevokeCacheSecurityGroupIngress
  Action: elasticache:RevokeCacheSecurityGroupIngress
Resource: *
Compliance Validation for Amazon ElastiCache

Third-party auditors assess the security and compliance of Amazon ElastiCache as part of multiple AWS compliance programs. These include SOC, PCI, FedRAMP, HIPAA, and others.

For a list of AWS services in scope of specific compliance programs, see AWS Services in Scope by Compliance Program. For general information, see AWS Compliance Programs.

You can download third-party audit reports using AWS Artifact. For more information, see Downloading Reports in AWS Artifact.

Your compliance responsibility when using Amazon ElastiCache is determined by the sensitivity of your data, your company's compliance objectives, and applicable laws and regulations. If your use of Amazon ElastiCache is subject to compliance with standards such as HIPAA, PCI, or FedRAMP, AWS provides resources to help:

- **Security and Compliance Quick Start Guides** – These deployment guides discuss architectural considerations and provide steps for deploying security- and compliance-focused baseline environments on AWS.
- **Architecting for HIPAA Security and Compliance Whitepaper** – This whitepaper describes how companies can use AWS to create HIPAA-eligible applications.
- **AWS Compliance Resources** – This collection of workbooks and guides might apply to your industry and location.
- **AWS Config** – This AWS service assesses how well your resource configurations comply with internal practices, industry guidelines, and regulations.
- **AWS Security Hub** – This AWS service provides a comprehensive view of your security state within AWS that helps you check your compliance with security industry standards and best practices.

Resilience in Amazon ElastiCache

The AWS global infrastructure is built around AWS Regions and Availability Zones. AWS Regions provide multiple physically separated and isolated Availability Zones, which are connected with low-latency, high-throughput, and highly redundant networking. With Availability Zones, you can design and operate applications and databases that automatically fail over between Availability Zones without interruption. Availability Zones are more highly available, fault tolerant, and scalable than traditional single or multiple data center infrastructures.

For more information about AWS Regions and Availability Zones, see AWS Global Infrastructure.

In addition to the AWS global infrastructure, Amazon ElastiCache offers several features to help support your data resiliency and backup needs.

**Topics**

- **Mitigating Failures (p. 251)**

Mitigating Failures

When planning your Amazon ElastiCache implementation, you should plan so that failures have a minimal impact upon your application and data. The topics in this section cover approaches you can take to protect your application and data from failures.

**Topics**
Mitigating Failures when Running Memcached

When running the Memcached engine, you have the following options for minimizing the impact of a failure. There are two types of failures to address in your failure mitigation plans: node failure and Availability Zone failure.

Mitigating Node Failures

To mitigate the impact of a node failure, spread your cached data over more nodes. Because Memcached does not support replication, a node failure will always result in some data loss from your cluster.

When you create your Memcached cluster you can create it with 1 to 20 nodes, or more by special request. Partitioning your data across a greater number of nodes means you'll lose less data if a node fails. For example, if you partition your data across 10 nodes, any single node stores approximately 10% of your cached data. In this case, a node failure loses approximately 10% of your cache which needs to be replaced when a replacement node is created and provisioned. If the same data were cached in 3 larger nodes, the failure of a node would lose approximately 33% of your cached data.

If you need more than 20 nodes in a Memcached cluster, or more than 100 nodes total in an AWS Region, fill out the ElastiCache Limit Increase Request form at https://aws.amazon.com/contact-us/elasticache-node-limit-request/.

For information on specifying the number of nodes in a Memcached cluster, see Creating a Memcached Cluster (Console) (p. 82).

Mitigating Availability Zone Failures

To mitigate the impact of an Availability Zone failure, locate your nodes in as many Availability Zones as possible. In the unlikely event of an AZ failure, you will lose the data cached in that AZ, not the data cached in the other AZs.

Why so many nodes?

If my region has only 3 Availability Zones, why do I need more than 3 nodes since if an AZ fails I lose approximately one-third of my data?

This is an excellent question. Remember that we’re attempting to mitigate two distinct types of failures, node and Availability Zone. You’re right, if your data is spread across Availability Zones and one of the zones fails, you will lose only the data cached in that AZ, irrespective of the number of nodes you have. However, if a node fails, having more nodes will reduce the proportion of data lost.

There is no "magic formula" for determining how many nodes to have in your cluster. You must weight the impact of data loss vs. the likelihood of a failure vs. cost, and come to your own conclusion.

For information on specifying the number of nodes in a Memcached cluster, see Creating a Memcached Cluster (Console) (p. 82).

For more information on regions and Availability Zones, see Choosing Regions and Availability Zones (p. 37).

Recommendations

There are two types of failures you need to plan for, individual node failures and broad Availability Zone failures. The best failure mitigation plan will address both kinds of failures.
Minimizing the Impact of Failures

To minimize the impact of a node failure, we recommend that your implementation use multiple nodes in each shard and distribute the nodes across multiple Availability Zones.

When running Memcached and partitioning your data across nodes, the more nodes you use the smaller the data loss if any one node fails.

Minimizing the Impact of Availability Zone Failures

To minimize the impact of an Availability Zone failure, we recommend launching your nodes in as many different Availability Zones as are available. Spreading your nodes evenly across AZs will minimize the impact in the unlikely event of an AZ failure.

Infrastructure Security in AWS Elasticache

As a managed service, AWS Elasticache is protected by the AWS global network security procedures that are described in the Amazon Web Services: Overview of Security Processes whitepaper.

You use AWS published API calls to access Elasticache through the network. Clients must support Transport Layer Security (TLS) 1.0 or later. We recommend TLS 1.2 or later. Clients must also support cipher suites with perfect forward secrecy (PFS) such as Ephemeral Diffie-Hellman (DHE) or Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). Most modern systems such as Java 7 and later support these modes.

Additionally, requests must be signed by using an access key ID and a secret access key that is associated with an IAM principal. Or you can use the AWS Security Token Service (AWS STS) to generate temporary security credentials to sign requests.
Managing the Service Updates

Amazon ElastiCache automatically monitors your fleet of Memcached clusters and nodes to apply service updates as they become available. This typically means that you must set up a pre-defined maintenance window so that ElastiCache can apply these updates, which you might find too rigid and might constrain your business flows. With self-service updates, you control when and which updates are applied. You can also monitor the progress of these updates to your selected Memcached clusters in real-time.

Depending on your business requirements, you can choose to stop the update to remaining nodes and cache clusters. You can select a new set of clusters (including the ones that were partially updated) to apply the service updates anytime until the service update expires.

Managing the Service Updates

ElastiCache service updates are released on a regular basis. If you have one or more qualifying cache clusters for those service updates, you receive notifications through email, SNS, the Personal Health Dashboard (PHD), and Amazon CloudWatch events. The updates are also displayed on the Service Updates page on the ElastiCache console.

However, you control when to apply an update, regardless of the recommendation. At a minimum, we strongly recommend that you apply any updates of type security to ensure that your Memcached clusters are always up-to-date with current security patches.

For more information, see Amazon ElastiCache Maintenance Help Page.

The following sections explore these options in detail.

Topics

- Applying the Self-Service Updates (p. 254)
- Stopping the Self-Service Updates (p. 259)

Applying the Self-Service Updates

You can start applying the service updates to your Memcached fleet from the time that the updates have an available status until they have an expired status. Service updates of the type security are cumulative. In other words, any nonexpired updates that you haven't applied yet are included with your latest update.

Note

You can apply only those service updates that have an available status, even if the recommended apply by date is past due.

For more information about reviewing your Memcached fleet and applying any service-specific updates to applicable Memcached clusters, see Applying the Service Updates Using the Console for Memcached (p. 255).
When a new service update is available for one or more Memcached clusters in your fleet, you can use the ElastiCache console, API, or AWS CLI to apply the update. The following sections explain the options that you can use to apply updates.

**Topics**
- Applying the Service Updates Using the Console (p. 255)
- Applying the Service Updates Using the AWS CLI (p. 259)

### Applying the Service Updates Using the Console

You can apply the service updates using one of the following console options. ElastiCache provides you two different perspectives to help you decide how and when to apply the updates:

**Topics**
- Applying the Service Updates Using the Console for Memcached (p. 255)
- Applying the Service Updates Using the Service Updates List (p. 257)

### Applying the Service Updates Using the Console for Memcached

Choose this to review the **Update Status** of individual Memcached clusters, and then choose **Apply**, **View**, or **Stop** for the service updates. If a service update is available, the console displays a banner at the top of the **Memcached** page, as shown following:

- If you choose **Apply Now**, you can choose to apply the service update to all or a subset of the applicable clusters in this workflow, as shown following:

  **Note**
  
  If you choose **Dismiss**, the console stops displaying the banner for that console session. However, the banner reappears the next time that you refresh your session.
Be aware of the following about the Apply Updates Now page:

- **Auto-Update after Due Date**: If you choose not to apply the self-service update before it expires, any clusters or individual nodes that aren’t updated remain out of compliance until the next cumulative update is available. ElastiCache doesn’t automatically apply the service update on your behalf.

- The ratio of **Nodes Updated** on your Memcached cluster and the **Estimated Update Time** allow you to plan your maintenance schedule. If service updates exceed the estimated time constraints for your business flows, you have the option to stop them and re-apply them at a later date. For more information, see Stopping the Self-Service Updates (p. 259).

- If you choose to apply the service updates to any or all available Memcached clusters, choose **Confirm**. If you choose this, you can then view the Service Updates page, where you can monitor the status of your service update.

- If you choose **Cancel**, you can explore further options, as explained following:

You can inspect your Memcached clusters on an individual basis to determine their **Update Status**. The following lets you know the compliance status of your clusters with regard to available service updates.
**Update Status** displays one of the following:

- **update available**: An update is available to apply to this cluster.
- **in-progress**: The update is being applied to this cluster, rendering it unavailable for business flows.
- **stopping**: An in-progress update has been interrupted before completion.
- **stopped**: The update has been terminated.

**Note**

If you stop an in-progress update on a Memcached cluster, some nodes might be updated while others are not. The **stopping** process doesn't roll back any changes to already updated nodes. You can re-apply the update to those nodes that still have an **available** status at your convenience, as long as the update doesn't have an **Expired** status.

- **up to date**: The update has been applied and your cluster is up to date.

**Applying the Service Updates Using the Service Updates List**

To review the list of individual service updates and their status, along with other relevant information, choose the **Service Updates List** tab.

When viewing the **Service Updates List**, note the following:

- **Service Update Name**: A unique identifier for the service update.
- **Status**: The status of the update, which will be one of the following:
  - **available**: The update is available for requisite Memcached clusters.
  - **complete**: The update has been applied and all Memcached clusters are up to date.
Applying the Self-Service Updates

- **cancelled**: The update has been cancelled and is no longer necessary.
- **expired**: The update is no longer available to apply.

**Severity**: The priority of applying the update:
- **critical**: Recommended to apply immediately (within 14 days or less).
- **important**: Recommended to apply as soon as your business flow allows (within 30 days or less).
- **medium**: Recommended to apply as soon as possible as your business flow allows (within 60 days or less).
- **low**: Recommended to apply as soon as possible as your business flow allows (within 90 days or less).

**Update Type**: For this version, only security updates are supported.

**Release Date**: When the update is released and available to apply on your Redis fleet.

**Recommended Apply By Date**: ElastiCache guidance date to apply the updates by.

Choosing an individual update provides additional details, including the following:

- **Update Description**: Provides details on the service update.
- **Update Expiration Date**: The date when the service update expires and no longer is available. Any updates that aren't applied before their expiration date are cumulatively rolled into the next update.

To review the list of individual service updates in relation to the applicable Memcached clusters, choose the Service Update Status tab.

When viewing the Service Updates Status list, note the following:

- **Service Update Name**: Provides detailed information about the service update.
- **Cluster Name**: The list of your Memcached clusters that are eligible for the update.
- **Nodes Updated**: The ratio of individual nodes within a specific cluster that were updated or remain available for the specific service update.
- **Update Severity**: The priority of applying the update:
  - **critical**: Recommended to apply immediately (within 14 days or less).
  - **important**: Recommended to apply as soon as your business flow allows (within 30 days or less).
  - **medium**: Recommended to apply as soon as possible as your business flow allows (within 60 days or less).
  - **low**: Recommended to apply as soon as possible as your business flow allows (within 90 days or less).
• **Update Type**: For this version, only security updates are supported.

• **Service Update Status**: The status of the update, which will be one of the following:
  • available: The update is available for requisite Memcached clusters.
  • complete: The update has been applied and all Memcached clusters are updated.
  • canceled: The update has been canceled and is no longer necessary.
  • expired: The update is no longer available to apply.

• **Service Update SLA Met**: This reflects whether your cluster is compliant.
  • yes: All available updates have been applied to this cluster and available nodes.
  • no: The service update might have been applied successfully to one or more nodes, but other nodes within the cluster still have an available status. This typically happens when a service update is applied and then stopped.

  **Note**
  If you stop the progress of a service update on a cluster, any nodes that are already updated have a complete status. Any nodes that have an In Progress or Stopping status revert to a Stopped status, and the Service Update SLA Met status changes to no.

• **Cluster Status Modified Date**: The latest date that the cluster was modified with a service update.

  **Note**
  The Show Previous Updates check box, if selected, displays a list of previous updates that are no longer available.

### Applying the Service Updates Using the AWS CLI

After you receive notification that service updates are available, you can inspect and apply them using the AWS CLI:

• To retrieve a description of the service updates that are available:

  ```
  aws elasticache describe-service-updates --service-update-status available
  ```

  For more information, see DescribeServiceUpdates.

• To review update actions that have a not-applied or stopped status:

  ```
  aws elasticache describe-update-actions --service-update-name sample-service-update --update-action-status not-applied stopped
  ```

  For more information, see DescribeUpdateActions.

• To apply a service update on a list of replication groups:

  ```
  ```

  For more information, see BatchApplyUpdateAction.

### Stopping the Self-Service Updates

If you have an unexpected surge to your Redis clusters that are undergoing updates, or if updates are taking too long and interrupting your business flow peak time, you can stop them.

The Stopping operation immediately interrupts all updates to those clusters and any nodes that are yet to be updated. It continues to completion any nodes that have an in progress status. However, it ceases updates to other nodes in the same cluster that have an update available status and reverts them to a Stopping status.
When the Stopping workflow is complete, the nodes that have a Stopping status change to a Stopped status. Depending on the workflow of the update, some clusters won't have any nodes updated. Other clusters might include some nodes that are updated and others that still have an update available status.

You can return later to finish the update process as your business flows permit. In this case, choose the applicable clusters that you want to complete updates on, and then choose Apply Now. For more information, see Applying the Service Updates Using the Console for Memcached (p. 255).

**Topics**

- Stopping the Service Updates Using the Console (p. 260)
- Stopping the Service Updates Using the AWS CLI (p. 261)

**Stopping the Service Updates Using the Console**

You can interrupt a service update using the console. The following demonstrates how to do this:

- After a service update has progressed on a selected Memcached cluster, the ElastiCache console displays the **View/Stop Update** tab at the top of the dashboard.

  ![ElastiCache Dashboard](image)

- To interrupt the update, choose **Stop Update**.
Stopping the Self-Service Updates

• When you stop the update, choose the Memcached cluster and examine the status. It will revert to a **Stopping** status and eventually a **Stopped** status.

Stopping the Service Updates Using the AWS CLI

You can interrupt a service update using the AWS CLI. The following code example shows how to do this:

```
aws elasticache batch-stop-update-action --service-update-name sample-service-update --cache-group-ids my-cache-group-1 my-cache-group-2
```

For more information, see BatchStopUpdateAction.
Using the ElastiCache API

This section provides task-oriented descriptions of how to use and implement ElastiCache operations. For a complete description of these operations, see the Amazon ElastiCache API Reference

Topics

- Using the Query API (p. 262)
- Available Libraries (p. 264)
- Troubleshooting Applications (p. 265)
- Logging Amazon ElastiCache API Calls with AWS CloudTrail (p. 265)

Using the Query API

Query Parameters

HTTP Query-based requests are HTTP requests that use the HTTP verb GET or POST and a Query parameter named Action.

Each Query request must include some common parameters to handle authentication and selection of an action.

Some operations take lists of parameters. These lists are specified using the param.n notation. Values of n are integers starting from 1.

Query Request Authentication

You can only send Query requests over HTTPS and you must include a signature in every Query request. This section describes how to create the signature. The method described in the following procedure is known as signature version 4.

The following are the basic steps used to authenticate requests to AWS. This assumes you are registered with AWS and have an Access Key ID and Secret Access Key.

Query Authentication Process

1. The sender constructs a request to AWS.
2. The sender calculates the request signature, a Keyed-Hashing for Hash-based Message Authentication Code (HMAC) with a SHA-1 hash function, as defined in the next section of this topic.
3. The sender of the request sends the request data, the signature, and Access Key ID (the key-identifier of the Secret Access Key used) to AWS.
4. AWS uses the Access Key ID to look up the Secret Access Key.
5. AWS generates a signature from the request data and the Secret Access Key using the same algorithm used to calculate the signature in the request.
6. If the signatures match, the request is considered to be authentic. If the comparison fails, the request is discarded, and AWS returns an error response.

**Note**
If a request contains a Timestamp parameter, the signature calculated for the request expires 15 minutes after its value.
If a request contains an Expires parameter, the signature expires at the time specified by the Expires parameter.

**To calculate the request signature**

1. Create the canonicalized query string that you need later in this procedure:
   a. Sort the UTF-8 query string components by parameter name with natural byte ordering. The parameters can come from the GET URI or from the POST body (when Content-Type is application/x-www-form-urlencoded).
   b. URL encode the parameter name and values according to the following rules:
      i. Do not URL encode any of the unreserved characters that RFC 3986 defines. These unreserved characters are A-Z, a-z, 0-9, hyphen (-), underscore (_), period (.), and tilde (~).
      ii. Percent encode all other characters with %XY, where X and Y are hex characters 0-9 and uppercase A-F.
      iii. Percent encode extended UTF-8 characters in the form %XY%ZA....
      iv. Percent encode the space character as %20 (and not +, as common encoding schemes do).
   c. Separate the encoded parameter names from their encoded values with the equals sign (=) (ASCII character 61), even if the parameter value is empty.
   d. Separate the name-value pairs with an ampersand (&) (ASCII code 38).
2. Create the string to sign according to the following pseudo-grammar (the "\n" represents an ASCII newline).

   ```
   StringToSign = HTTPVerb + "\n" +
   ValueOfHostHeaderInLowercase + "\n" +
   HTTPRequestURI + "\n" +
   CanonicalizedQueryString <from the preceding step>
   ```

   The HTTPRequestURI component is the HTTP absolute path component of the URI up to, but not including, the query string. If the HTTPRequestURI is empty, use a forward slash (/).
3. Calculate an RFC 2104-compliant HMAC with the string you just created, your Secret Access Key as the key, and SHA256 or SHA1 as the hash algorithm.
   For more information, see https://www.ietf.org/rfc/rfc2104.txt.
4. Convert the resulting value to base64.
5. Include the value as the value of the Signature parameter in the request.

For example, the following is a sample request (linebreaks added for clarity).

```https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheClusters
```
For the preceding query string, you would calculate the HMAC signature over the following string.

```
GET

elasticache.amazonaws.com
Action=DescribeCacheClusters
&CacheClusterIdentifier=myCacheCluster
&SignatureMethod=HmacSHA256
&SignatureVersion=4
&Version=2014-12-01
&X-Amz-Algorithm=AWS4-HMAC-SHA256
&X-Amz-Credential=AKIADQKE4SARGYLE%2F20140523%2Fus-west-2%2Felasticache%2Faws4_request
&X-Amz-Date=20141201T223649Z
&X-Amz-SignedHeaders=content-type;host;user-agent;x-amz-content-sha256;x-amz-date
content-type:application/x-amz-json-1.1
host:elasticache.us-west-2.amazonaws.com
user-agent:CacheServicesAPIClient
x-amz-content-sha256:ZjP13zw9loMxd65n9yUUnvqJX1uX88x9AvmtiSKQ7cQ=
x-amz-date:20141201T223649Z

The result is the following signed request.

https://elasticache.us-west-2.amazonaws.com/

?Action=DescribeCacheClusters
&CacheClusterIdentifier=myCacheCluster
&SignatureMethod=HmacSHA256
&SignatureVersion=4
&Version=2014-12-01
&X-Amz-Algorithm=AWS4-HMAC-SHA256
&X-Amz-Credential=AKIADQKE4SARGYLE/20141201/us-west-2/elasticache/aws4_request
&X-Amz-Date=20141201T223649Z
&X-Amz-SignedHeaders=content-type;host;user-agent;x-amz-content-sha256;x-amz-date
&X-Amz-Signature=2877960fced9040b41b4feaca835f5d5cfe4b9264f768e6a0236c9145f915ffaf56
```

For detailed information on the signing process and calculating the request signature, see the topic [Signature Version 4 Signing Process and its subtopics](#).

**Available Libraries**

AWS provides software development kits (SDKs) for software developers who prefer to build applications using language-specific APIs instead of the Query API. These SDKs provide basic functions (not included in the APIs), such as request authentication, request retries, and error handling so that it is easier to get started. SDKs and additional resources are available for the following programming languages:

- **Java**
- **Windows and .NET**
- **PHP**
- **Python**
- **Ruby**

For information about other languages, see [Sample Code & Libraries](#).
Amazon ElastiCache ElastiCache
for Memcached User Guide
Troubleshooting Applications

Troubleshooting Applications
ElastiCache provides speciﬁc and descriptive errors to help you troubleshoot problems while interacting
with the ElastiCache API.

Retrieving Errors
Typically, you want your application to check whether a request generated an error before you spend any
time processing results. The easiest way to ﬁnd out if an error occurred is to look for an Error node in
the response from the ElastiCache API.
XPath syntax provides a simple way to search for the presence of an Error node, as well as an easy way
to retrieve the error code and message. The following code snippet uses Perl and the XML::XPath module
to determine if an error occurred during a request. If an error occurred, the code prints the ﬁrst error
code and message in the response.
use XML::XPath;
my $xp = XML::XPath->new(xml =>$response);
if ( $xp->find("//Error") )
{print "There was an error processing your request:\n", " Error code: ",
$xp->findvalue("//Error[1]/Code"), "\n", " ",
$xp->findvalue("//Error[1]/Message"), "\n\n"; }

Troubleshooting Tips
We recommend the following processes to diagnose and resolve problems with the ElastiCache API.
• Verify that ElastiCache is running correctly.
To do this, simply open a browser window and submit a query request to the ElastiCache service (such
as https://elasticache.amazonaws.com). A MissingAuthenticationTokenException or 500 Internal
Server Error conﬁrms that the service is available and responding to requests.
• Check the structure of your request.
Each ElastiCache operation has a reference page in the ElastiCache API Reference. Double-check that
you are using parameters correctly. To give you ideas regarding what might be wrong, look at the
sample requests or user scenarios to see if those examples are doing similar operations.
• Check the forum.
ElastiCache has a discussion forum where you can search for solutions to problems others have
experienced along the way. To view the forum, see

Logging Amazon ElastiCache API Calls with AWS
CloudTrail
Amazon ElastiCache is integrated with AWS CloudTrail, a service that provides a record of actions
taken by a user, role, or an AWS service in Amazon ElastiCache. CloudTrail captures all API calls for
Amazon ElastiCache as events, including calls from the Amazon ElastiCache console and from code
calls to the Amazon ElastiCache API operations. If you create a trail, you can enable continuous delivery
of CloudTrail events to an Amazon S3 bucket, including events for Amazon ElastiCache. 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
API Version 2015-02-02
265


Amazon ElastiCache for Memcached User Guide

Logging Amazon ElastiCache API Calls with AWS CloudTrail

ElastiCache, 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 ElastiCache Information in CloudTrail

CloudTrail is enabled on your AWS account when you create the account. When activity occurs in Amazon ElastiCache, 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 ElastiCache, 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 the following:

- 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 ElastiCache actions are logged by CloudTrail and are documented in the ElastiCache API Reference. For example, calls to the CreateCacheCluster, DescribeCacheCluster and ModifyCacheCluster 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 ElastiCache 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 example shows a CloudTrail log entry that demonstrates the CreateCacheCluster action.

```json
{
    "eventVersion":"1.01",
    "userIdentity":{
        "type":"IAMUser",
        "principalId":"EXAMPLEEXAMPLEEXAMPLE",
        "arn":"arn:aws:iam::123456789012:user/elasticache-allow",
        "accountId":"123456789012",
```
The following example shows a CloudTrail log entry that demonstrates the DescribeCacheCluster action. Note that for all Amazon ElastiCache Describe calls (Describe*), the ResponseElements section is removed and appears as null.

```json
{
    "eventVersion": "1.01",
    "userIdentity": {
        "type": "IAMUser",
        "principalId": "EXAMPLEEXAMPLEEXAMPLE",
        "arn": "arn:aws:iam::123456789012:user/elasticache-allow",
        "accountId": "123456789012",
        "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
        "userName": "elasticache-allow"
    },
    "eventTime": "2014-12-01T01:00:02Z",
    "eventSource": "elasticache.amazonaws.com",
    "eventName": "DescribeCacheClusters",
    " awsRegion": "us-west-2",
    "sourceIPAddress": "192.0.2.01",
    "userName": "elasticache-allow"
}
```

The example shows a CloudTrail log entry for the DescribeCacheClusters API call. Note that for all Amazon ElastiCache Describe calls (Describe*), the ResponseElements section is removed and appears as null.
The following example shows a CloudTrail log entry that records a **ModifyCacheCluster** action.

```json
{
    "eventVersion": "1.01",
    "userIdentity": {
        "type": "IAMUser",
        "principalId": "EXAMPLEEXAMPLEEXAMPLE",
        "arn": "arn:aws:iam::123456789012:user/elasticache-allow",
        "accountId": "123456789012",
        "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
        "userName": "elasticache-allow"
    },
    "eventTime": "2014-12-01T22:32:21Z",
    "eventSource": "elasticache.amazonaws.com",
    "eventName": "ModifyCacheCluster",
    "awsRegion": "us-west-2",
    "sourceIPAddress": "192.0.2.01",
    "userAgent": "Amazon CLI/ElastiCache 1.10 API 2014-12-01",
    "requestParameters": {
        "applyImmediately": true,
        "numCacheNodes": 3,
        "cacheClusterId": "test-memcached"
    },
    "responseElements": {
        "engine": "memcached",
        "clientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download",
        "cacheParameterGroup": {
            "cacheParameterGroupName": "default.memcached1.4",
            "cacheNodeIdsToReboot": [],
            "parameterApplyStatus": "in-sync"
        },
        "cacheClusterCreateTime": "Dec 1, 2014 10:16:06 PM",
        "preferredAvailabilityZone": "Multiple",
        "numCacheNodes": 2,
        "cacheNodeType": "cache.m1.small",
        "cacheClusterStatus": "modifying",
        "autoMinorVersionUpgrade": true,
        "preferredMaintenanceWindow": "thu:05:00-thu:06:00",
        "cacheClusterId": "test-memcached",
        "engineVersion": "1.4.14",
        "cacheSecurityGroups": [
            {
                "status": "active",
                "cacheSecurityGroupName": "default"
            }
        ],
        "configurationEndpoint": {
            "address": "test-memcached.example.cfg.use1prod.cache.amazonaws.com",
            "port": 11211
        },
        "pendingModifiedValues": {
            "showCacheNodeInfo": false,
            "maxRecords": 100
        },
        "requestID": "1f0b5031-3548-11e4-9376-c1d979ba565a",
        "eventID": "a58572a8-e81b-4100-8e00-1797ed19d172"
    }
}
Setting Up the ElastiCache Command Line Interface

This section describes the prerequisites for running the command line tools, where to get the command line tools, how to set up the tools and their environment, and includes a series of common examples of tool usage.

Follow the instructions in this topic only if you are going to the AWS CLI for ElastiCache.

**Important**
The Amazon ElastiCache Command Line Interface (CLI) does not support any ElastiCache improvements after API version 2014-09-30. To use newer ElastiCache functionality from the command line, use the AWS Command Line Interface.

**Topics**
- Prerequisites (p. 269)
- Getting the Command Line Tools (p. 270)
- Setting Up the Tools (p. 270)
- Providing Credentials for the Tools (p. 271)
- Environmental Variables (p. 272)

**Prerequisites**

This document assumes that you can work in a Linux/UNIX or Windows environment. The Amazon ElastiCache command line tools also work on Mac OS X, which is a UNIX-based environment; however, no specific Mac OS X instructions are included in this guide.

As a convention, all command line text is prefixed with a generic `PROMPT>` command line prompt. The actual command line prompt on your machine is likely to be different. We also use `$` to indicate a Linux/UNIX specific command and `C:\>` for a Windows specific command. The example output resulting from the command is shown immediately thereafter without any prefix.

**The Java Runtime Environment**

The command line tools used in this guide require Java version 5 or later to run. Either a JRE or JDK installation is acceptable. To view and download JREs for a range of platforms, including Linux/UNIX and Windows, see Java SE Downloads.

**Setting the Java Home Variable**

The command line tools depend on an environment variable (JAVA_HOME) to locate the Java Runtime. This environment variable should be set to the full path of the directory that contains a subdirectory named bin which in turn contains the executable java (on Linux and UNIX) or java.exe (on Windows) executable.

To set the Java Home variable
1. Set the Java Home variable.
   - On Linux and UNIX, enter the following command:

   ```
   $ export JAVA_HOME=<PATH>
   ```

   - On Windows, enter the following command:

   ```
   C:\> set JAVA_HOME=<PATH>
   ```

2. Confirm the path setting by running `$JAVA_HOME/bin/java -version` and checking the output.
   - On Linux/UNIX, you will see output similar to the following:

   ```
   $ $JAVA_HOME/bin/java -version
   java version "1.6.0_23"
   Java(TM) SE Runtime Environment (build 1.6.0_23-b05)
   Java HotSpot(TM) Client VM (build 19.0-b09, mixed mode, sharing)
   ```

   - On Windows, you will see output similar to the following:

   ```
   C:\> %JAVA_HOME%\bin\java -version
   java version "1.6.0_23"
   Java(TM) SE Runtime Environment (build 1.6.0_23-b05)
   Java HotSpot(TM) Client VM (build 19.0-b09, mixed mode, sharing)
   ```

## Getting the Command Line Tools

The command line tools are available as a ZIP file on the ElastiCache Developer Tools web site. These tools are written in Java, and include shell scripts for Windows 2000/XP/Vista/Windows 7, Linux/UNIX, and Mac OSX. The ZIP file is self-contained and no installation is required; simply download the zip file and unzip it to a directory on your local machine.

## Setting Up the Tools

The command line tools depend on an environment variable (AWS_ELASTICACHE_HOME) to locate supporting libraries. You need to set this environment variable before you can use the tools. Set it to the path of the directory you unzipped the command line tools into. This directory is named ElastiCacheCli-A.B.nnnn (A, B and n are version/release numbers), and contains subdirectories named bin and lib.

To set the AWS_ELASTICACHE_HOME environment variable

- Open a command line window and enter one of the following commands to set the AWS_ELASTICACHE_HOME environment variable.

  - On Linux and UNIX, enter the following command:

    ```
    $ export AWS_ELASTICACHE_HOME=<path-to-tools>
    ```

  - On Windows, enter the following command:

    ```
    C:\> set AWS_ELASTICACHE_HOME=<path-to-tools>
    ```
To make the tools easier to use, we recommend that you add the tools’ BIN directory to your system PATH. The rest of this guide assumes that the BIN directory is in your system path.

**To add the tools' BIN directory to your system path**

- Enter the following commands to add the tools' BIN directory to your system PATH.

  - On Linux and UNIX, enter the following command:

    ```
    $ export PATH=$PATH:$AWS_ELASTICACHE_HOME/bin
    ```

  - On Windows, enter the following command:

    ```
    C:\> set PATH=%PATH%;%AWS_ELASTICACHE_HOME%/bin
    ```

**Note**
The Windows environment variables are reset when you close the command window. You might want to set them permanently. Consult the documentation for your version of Windows for more information.

**Note**
Paths that contain a space must be wrapped in double quotes, for example: "C:\Program Files\Java"

### Providing Credentials for the Tools

The command line tools need the AWS Access Key and Secret Access Key provided with your AWS account. You can get them using the command line or from a credential file located on your local system.

The deployment includes a template file `${AWS_ELASTICACHE_HOME}/credential-file-path.template` that you need to edit with your information. Following are the contents of the template file:

```
AWSAccessKeyId=<Write your AWS access ID>
AWSSecretKey=<Write your AWS secret key>
```

**Important**
On UNIX, limit permissions to the owner of the credential file:

```
$ chmod 600 <the file created above>
```

With the credentials file setup, you'll need to set the AWS_CREDENTIAL_FILE environment variable so that the ElastiCache tools can find your information.

**To set the AWS_CREDENTIAL_FILE environment variable**

1. Set the environment variable:

   - On Linux and UNIX, update the variable using the following command:

     ```
     $ export AWS_CREDENTIAL_FILE=<the file created above>
     ```

   - On Windows, set the variable using the following command:

     ```
     C:\> set AWS_CREDENTIAL_FILE=<the file created above>
     ```
2. Check that your setup works properly, run the following command:

```
elasticache --help
```

You should see the usage page for all ElastiCache commands.

### Environmental Variables

Environment variables can be useful for scripting, configuring defaults or temporarily overriding them.

In addition to the AWS_CREDENTIAL_FILE environment variable, most API tools included with the ElastiCache Command Line Interface support the following variables:

- **EC2_REGION** — The AWS region to use.
- **AWS_ELASTICACHE_URL** — The URL to use for the service call. Not required to specify a different regional endpoint if EC2_REGION is specified or the --region parameter is passed.

The following examples show how to set the environmental variable EC2_REGION to configure the region used by the API tools:

**Linux, OS X, or Unix**

```
$ export EC2_REGION=us-west-1
```

**Windows**

```
$ set EC2_REGION=us-west-1
```

### Amazon ElastiCache Error Messages

The following error messages are returned by Amazon ElastiCache. You may receive other error messages that are returned by ElastiCache, other AWS services, or by Memcached. For descriptions of error messages from sources other than ElastiCache, see the documentation from the source that is generating the error message.

- **Cluster node quota exceeded (p. 272)**
- **Customer's node quota exceeded (p. 273)**
- **Insufficient cache cluster capacity (p. 273)**

Error Message: **Cluster node quota exceeded. Each cluster can have at most \( %n \) nodes in this region.**

**Cause:** You attempted to create or modify a cluster with the result that the cluster would have more than \( %n \) nodes.

**Solution:** Change your request so that the cluster does not have more than \( %n \) nodes. Or, if you need more than \( %n \) nodes, make your request using the Amazon ElastiCache Node request form.

For more information, see Amazon ElastiCache Limits in Amazon Web Services General Reference.
Error Messages: **Customer node quota exceeded. You can have at most %n nodes in this region** Or, **You have already reached your quota of %s nodes in this region.**

**Cause:** You attempted to create or modify a cluster with the result that your account would have more than %n nodes across all clusters in this region.

**Solution:** Change your request so that the total nodes in the region across all clusters for this account does not exceed %n. Or, if you need more than %n nodes, make your request using the Amazon ElastiCache Node request form.

For more information, see Amazon ElastiCache Limits in *Amazon Web Services General Reference*.

Error Messages: **InsufficientCacheClusterCapacity**

**Cause:** AWS does not currently have enough available On-Demand capacity to service your request.

**Solution:**
- Wait a few minutes and then submit your request again; capacity can shift frequently.
- Submit a new request with a reduced number of nodes or shards (node groups). For example, if you're making a single request to launch 15 nodes, try making 3 requests of 5 nodes, or 15 requests for 1 node instead.
- If you're launching a cluster, submit a new request without specifying an Availability Zone.
- If you're launching a cluster, submit a new request using a different node type (which you can scale up at a later stage). For more information, see Scaling ElastiCache for Memcached Clusters (p. 132).

Notifications

This topic covers ElastiCache notifications that you might be interested in. A notification is a situation or event that, in most cases, is temporary, lasting only until a solution is found and implemented. Notifications generally have a start date and a resolution date, after which the notification is no longer relevant. Any one notification might or might not be relevant to you. We recommend an implementation guideline that, if followed, improves the performance of your cluster.

Notifications do not announce new or improved ElastiCache features or functionality.

**General ElastiCache Notifications**

Currently there are no outstanding ElastiCache notifications that are not engine specific.

**ElastiCache for Memcached Notifications**

The following ElastiCache notifications are specific to the Memcached engine.

**ElastiCache for Memcached specific notifications**

- **Alert: Memcached LRU Crawler Causing Segmentation Faults** (p. 273)

**Alert: Memcached LRU Crawler Causing Segmentation Faults**

Alert Date: February 28, 2017
In some circumstances, your cluster might display instability with a segmentation fault in the Memcached LRU Crawler. This is an issue within the Memcached engine that has existed for some time. The issue became apparent in Memcached 1.4.33 when the LRU Crawler was enabled by default.

If you are experiencing this issue, we recommend that you disable the LRU Crawler until there is a fix. To do so, use `lru_crawler disable` at the command line or modify the `lru_crawler` parameter value (preferred).

Resolved Date:
Resolution:
Documentation History

- API version: 2015-02-02
- Latest documentation update: November 12, 2019

The following table describes important changes in each release of the *ElastiCache for Memcached User Guide* after March 2018. For notification about updates to this documentation, you can subscribe to the RSS feed.

**Recent ElastiCache for Memcached Updates**

<table>
<thead>
<tr>
<th>update-history-change</th>
<th>update-history-description</th>
<th>update-history-date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon ElastiCache now supports T3-Standard cache nodes (p. 275)</td>
<td>You can now launch the next generation general-purpose burstable T3-Standard cache nodes in Amazon ElastiCache. Amazon EC2's T3-Standard instances provide a baseline level of CPU performance with the ability to burst CPU usage at any time until the accrued credits are exhausted. For more information, see Supported Node Types.</td>
<td>November 12, 2019</td>
</tr>
<tr>
<td>ElastiCache for Memcached now allows users to apply service updates on their own schedule (p. 275)</td>
<td>With this feature, you can choose to apply available service updates at a time of your choosing and not just during maintenance windows. This will minimize service interruptions, particularly during peak business flows, and help ensure you remain compliant with security updates. For more information, see Self-Service Updates in Amazon ElastiCache.</td>
<td>October 9, 2019</td>
</tr>
<tr>
<td>Support for ElastiCache for Memcached 1.5.16</td>
<td>ElastiCache for Memcached now supports Memcached 1.5.16. It includes cumulative bug fixes from versions Memcached 1.5.14 and Memcached 1.5.15. For more information, see Memcached Version 1.5.16.</td>
<td>September 6, 2019</td>
</tr>
<tr>
<td>ElastiCache Standard Reserved Instance offerings: Partial Upfront, All Upfront and No Upfront. (p. 275)</td>
<td>Reserved Instances give you the flexibility to reserve an Amazon ElastiCache instance for a one- or three-year term based on an instance type and AWS Region. For more information, see Managing Costs with Reserved Nodes.</td>
<td>January 18, 2019</td>
</tr>
</tbody>
</table>
ElastiCache for Memcached 1.5.10

Support for Asia Pacific (Osaka-Local).

ElastiCache added support for the Asia Pacific (Osaka-Local) Region. The Asia Pacific (Osaka-Local) Region currently supports a single Availability Zone and is by invitation only. For more information, see the following:

- Supported Regions & Endpoints (p. 38)
- Supported Node Types (p. 46)

Support for EU (Paris).

ElastiCache added support for the EU (Paris) Region. For more information, see the following:

- Supported Regions & Endpoints (p. 38)
- Supported Node Types (p. 46)

Support for China (Ningxia) Region

Amazon ElastiCache added support for China (Ningxia) Region. For more information, see the following:

- Supported Regions & Endpoints (p. 38)
- Supported Node Types (p. 46)

Support for Service Linked Roles

This release of ElastiCache added support for Service Linked Roles (SLR). For more information, see the following:

- Using Service-Linked Roles for Amazon ElastiCache (p. 241)

The following table describes the important changes to the *ElastiCache for Memcached User Guide* before March 2018.

<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support for Asia Pacific (Osaka-Local)</td>
<td>ElastiCache added support for the Asia Pacific (Osaka-Local) Region. The Asia Pacific (Osaka-Local) Region currently supports a single Availability Zone and is by invitation only. For more information, see the following:</td>
<td>February 12, 2018</td>
</tr>
<tr>
<td></td>
<td>• Supported Regions &amp; Endpoints (p. 38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Supported Node Types (p. 46)</td>
<td></td>
</tr>
<tr>
<td>Support for EU (Paris)</td>
<td>ElastiCache added support for the EU (Paris) Region. For more information, see the following:</td>
<td>December 18, 2017</td>
</tr>
<tr>
<td></td>
<td>• Supported Regions &amp; Endpoints (p. 38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Supported Node Types (p. 46)</td>
<td></td>
</tr>
<tr>
<td>Support for China (Ningxia) Region</td>
<td>Amazon ElastiCache added support for China (Ningxia) Region. For more information, see the following:</td>
<td>December 11, 2017</td>
</tr>
<tr>
<td></td>
<td>• Supported Regions &amp; Endpoints (p. 38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Supported Node Types (p. 46)</td>
<td></td>
</tr>
<tr>
<td>Support for Service Linked Roles</td>
<td>This release of ElastiCache added support for Service Linked Roles (SLR). For more information, see the following:</td>
<td>December 7, 2017</td>
</tr>
<tr>
<td></td>
<td>• Using Service-Linked Roles for Amazon ElastiCache (p. 241)</td>
<td></td>
</tr>
<tr>
<td>Change</td>
<td>Description</td>
<td>Date Changed</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
</tbody>
</table>
| Support for R4 node types | This release of ElastiCache added support R4 node types in all AWS Regions supported by ElastiCache. You can purchase R4 node types as On-Demand or as Reserved Cache Nodes. For more information, see the following:  
- Supported Node Types (p. 46)  
- Memcached Node-Type Specific Parameters (p. 158) | November 20, 2017 |
| Connection patterns topic | ElastiCache documentation adds a topic covering various patterns for accessing an ElastiCache cluster in an Amazon VPC. For more information, see Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 204) in the ElastiCache User Guide. | April 24, 2017 |
| Support for Memcached 1.4.34 | ElastiCache adds support Memcached version 1.4.34, which incorporates a number of fixes to earlier Memcached versions. For more information, see Memcached 1.4.34 Release Notes at Memcached on GitHub. | April 10, 2017 |
| Support for Memcached 1.4.33 | ElastiCache adds support for Memcached version 1.4.33. For more information, see the following:  
- Memcached Version 1.4.33 (p. 35)  
- Memcached 1.4.33 Added Parameters (p. 152) | December 20, 2016 |
| Support for EU West (London) Region | ElastiCache adds support for EU (London) Region. Only node types T2 and M4 are currently supported. For more information, see the following:  
- Supported Regions & Endpoints (p. 38)  
- Supported Node Types (p. 46) | December 13, 2016 |
| Support for Canada (Montreal) Region | ElastiCache adds support for the Canada (Montreal) Region. Only node type M4 and T2 are currently supported in this AWS Region. For more information, see the following:  
- Supported Regions & Endpoints (p. 38)  
- Supported Node Types (p. 46) | December 8, 2016 |
| Support for M4 and R3 node types | ElastiCache adds support for R3 and M4 node types in South America (São Paulo) Region and M4 node types in China (Beijing) Region. For more information, see the following:  
- Supported Regions & Endpoints (p. 38)  
- Supported Node Types (p. 46) | November 1, 2016 |
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
</table>
| US East 2 (Ohio) Region support | ElastiCache adds support for the US East (Ohio) Region (us-east-2) with M4, T2, and R3 node types. For more information, see the following:  
   - Supported Regions & Endpoints (p. 38)  
   - Supported Node Types (p. 46) | October 17, 2016 |
| M4 node type support | ElastiCache adds support for the M4 family of node types in most AWS Regions supported by ElastiCache. You can purchase M4 node types as On-Demand or as Reserved Cache Nodes. For more information, see the following:  
   - Supported Node Types (p. 46)  
   - Memcached Node-Type Specific Parameters (p. 158) | August 3, 2016 |
| Mumbai Region support | ElastiCache adds support for the Asia Pacific (Mumbai) Region. For more information, see the following:  
   - Supported Node Types (p. 46)  
   - Memcached Node-Type Specific Parameters (p. 158) | June 27, 2016 |
| Support for R3 node types | ElastiCache adds support for R3 node types in the China (Beijing) Region and South America (São Paulo) Region. For more information, see Supported Node Types (p. 46). | March 16, 2016 |
| Accessing ElastiCache using a Lambda function | Added a tutorial on configuring a Lambda function to access ElastiCache in an Amazon VPC. For more information, see ElastiCache Tutorials and Videos (p. 23). | February 12, 2016 |
| Support for Asia Pacific (Seoul) Region | ElastiCache adds support for the Asia Pacific (Seoul) (ap-northeast-2) Region with t2, m3, and r3 node types. | January 6, 2016 |
| Support for Memcached 1.4.28. | ElastiCache adds support for Memcached version 1.4.24 and Memcached improvements since version 1.4.14. This release adds support for least recently used (LRU) cache management as a background task, choice of jenkins or murmur3 as your hashing algorithm, new commands, and miscellaneous bug fixes. For more information, see the following:  
   - Memcached release notes  
<p>| Support for Memcached Auto Discovery using PHP 5.6 | This release of Amazon ElastiCache adds support for Memcached Auto Discovery client for PHP version 5.6. For more information, see Compiling the Source Code for the ElastiCache Cluster Client for PHP (p. 77). | July 29, 2015 |</p>
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>New topic: Accessing ElastiCache from outside AWS</td>
<td>Added new topic on how to access ElastiCache resources from outside AWS. For more information, see Accessing ElastiCache Resources from Outside AWS (p. 111).</td>
<td>July 9, 2015</td>
</tr>
<tr>
<td>Node replacement messages added</td>
<td>ElastiCache adds three messages pertaining to scheduled node replacement, ElastiCache:NodeReplacementScheduled, ElastiCache:NodeReplacementRescheduled, and ElastiCache:NodeReplacementCanceled. For more information and actions you can take when a node is scheduled for replacement, see ElastiCache's Event Notifications and Amazon SNS (p. 175).</td>
<td>June 11, 2015</td>
</tr>
<tr>
<td>Support for cost allocation tags</td>
<td>ElastiCache adds support for cost allocation tags. For more information, see Monitoring Costs with Cost Allocation Tags (p. 179).</td>
<td>February 9, 2015</td>
</tr>
<tr>
<td>Support for Europe (Frankfurt) Region</td>
<td>ElastiCache adds support for the Europe (Frankfurt) (eu-central-1) Region.</td>
<td>January 19, 2015</td>
</tr>
<tr>
<td>AWS CloudTrail logging of API calls supported</td>
<td>ElastiCache adds support for using AWS CloudTrail to log all ElastiCache API calls. For more information, see Logging Amazon ElastiCache API Calls with AWS CloudTrail (p. 265).</td>
<td>September 15, 2014</td>
</tr>
<tr>
<td>New instance sizes supported</td>
<td>ElastiCache adds support for additional General Purpose (T2) instances. For more information, see Configuring Engine Parameters Using Parameter Groups (p. 135).</td>
<td>September 11, 2014</td>
</tr>
<tr>
<td>Flexible node placement supported for Memcached</td>
<td>ElastiCache adds support for creating Memcached nodes across multiple Availability Zones. For more information, see Step 1: Launch a Memcached Cluster (p. 17).</td>
<td>July 23, 2014</td>
</tr>
<tr>
<td>New instance sizes supported</td>
<td>ElastiCache adds support for additional General Purpose (M3) instances and Memory Optimized (R3) instances. For more information, see Configuring Engine Parameters Using Parameter Groups (p. 135).</td>
<td>July 1, 2014</td>
</tr>
<tr>
<td>PHP auto discovery</td>
<td>Added support for PHP version 5.5 auto discovery. For more information, see Installing the ElastiCache Cluster Client for PHP (p. 70).</td>
<td>May 13, 2014</td>
</tr>
<tr>
<td>Support for default Amazon Virtual Private Cloud (VPC)</td>
<td>In this release, ElastiCache is fully integrated with Amazon Virtual Private Cloud (VPC). For new customers, cache clusters are created in an Amazon VPC by default. For more information, see Amazon VPCs and ElastiCache Security (p. 198).</td>
<td>January 8, 2013</td>
</tr>
<tr>
<td>Change</td>
<td>Description</td>
<td>Date Changed</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>PHP support for cache node auto discovery</td>
<td>The initial release of cache node auto discovery provided support for Java programs. In this release, ElastiCache brings cache node auto discovery support to PHP.</td>
<td>January 2, 2013</td>
</tr>
<tr>
<td>Support for Amazon Virtual Private Cloud (VPC)</td>
<td>In this release, ElastiCache clusters can be launched in Amazon Virtual Private Cloud (VPC). By default, new customers’ cache clusters are created in an Amazon VPC automatically; existing customers can migrate to Amazon VPC at their own pace. For more information, see Amazon VPCs and ElastiCache Security (p. 198).</td>
<td>December 20, 2012</td>
</tr>
<tr>
<td>Cache node auto discovery and new cache engine version</td>
<td>ElastiCache provides cache node auto discovery—the ability for client programs to automatically determine all of the cache nodes in a cluster, and to initiate and maintain connections to all of these nodes. This release also offers a new cache engine version: Memcached version 1.4.14. This new cache engine provides enhanced slab rebalancing capability, significant performance and scalability improvements, and several bug fixes. There are several new cache parameters that can be configured. For more information, see Configuring Engine Parameters Using Parameter Groups (p. 135).</td>
<td>November 28, 2012</td>
</tr>
<tr>
<td>New cache node types</td>
<td>This release provides four additional cache node types.</td>
<td>November 13, 2012</td>
</tr>
<tr>
<td>Reserved cache nodes</td>
<td>This release adds support for reserved cache nodes.</td>
<td>April 5, 2012</td>
</tr>
<tr>
<td>New guide</td>
<td>This is the first release of Amazon ElastiCache User Guide.</td>
<td>August 22, 2011</td>
</tr>
</tbody>
</table>
AWS Glossary

For the latest AWS terminology, see the AWS Glossary in the AWS General Reference.