Amazon ElastiCache for Redis

ElastiCache for Redis User Guide

API Version 2015-02-02
# Table of Contents

What Is ElastiCache for Redis? ................................................................. 1
Common ElastiCache Use Cases and How ElastiCache Can Help ..................... 2
In-Memory Data Store ............................................................................. 2
Gaming Leaderboards (Redis Sorted Sets) .................................................. 4
Messaging (Redis Pub/Sub) ....................................................................... 5
Recommendation Data (Redis Hashes) ......................................................... 7
Other Redis Uses .................................................................................... 7
ElastiCache Customer Testimonials .......................................................... 7
ElastiCache Resources .............................................................................. 8
Components and Features .......................................................................... 10
Nodes ...................................................................................................... 10
ElastiCache for Redis Shards ..................................................................... 10
ElastiCache for Redis Clusters ................................................................... 11
ElastiCache for Redis Replication ............................................................. 12
AWS Regions and Availability Zones ......................................................... 13
ElastiCache for Redis Endpoints ............................................................... 13
Parameter Groups ................................................................................... 14
ElastiCache for Redis Security ................................................................. 14
Security Groups ..................................................................................... 14
Subnet Groups ....................................................................................... 15
ElastiCache for Redis Backups ................................................................. 15
Events ..................................................................................................... 15
ElastiCache for Redis Terminology .......................................................... 17
Tools for Managing Your Implementation .................................................. 20
Getting Started with ElastiCache for Redis ................................................ 24
Determine Requirements .......................................................................... 24
Setting Up ............................................................................................... 24
Create Your AWS Account ........................................................................ 24
Set Up Your Permissions (New ElastiCache Users Only) .............................. 25
Step 1: Launch a Cluster ........................................................................... 26
Step 2: Authorize Access ........................................................................ 28
Step 3: Connect to a Cluster's Node ........................................................... 29
Step 3.1: Find your Node Endpoints ......................................................... 29
Step 3.2: Connect to a Redis Cluster or Replication Group (Linux) ............... 29
Step 3.2: Connect to a Redis Cluster or Replication Group (Windows) .......... 31
Step 4: Delete Your Cluster ...................................................................... 32
Where Do I Go From Here? ..................................................................... 33
Tutorials and Videos ................................................................................. 34
Videos .................................................................................................... 35
Introductory Videos ................................................................................ 35
Advanced Videos .................................................................................... 35
Online Migration to ElastiCache ................................................................. 38
Overview ............................................................................................... 38
Migration Steps ..................................................................................... 38
Preparing Your Source and Target Redis Nodes for Migration ..................... 38
Starting Migration .................................................................................. 39
Verifying the Data Migration Progress ...................................................... 40
Completing the Data Migration ............................................................... 41
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing Your ElastiCache for Redis Implementation</td>
<td>44</td>
</tr>
<tr>
<td>Engine Versions and Upgrading</td>
<td>44</td>
</tr>
<tr>
<td>Supported Redis Versions</td>
<td>45</td>
</tr>
<tr>
<td>Upgrading Engine Versions</td>
<td>53</td>
</tr>
<tr>
<td>Choosing Regions and Availability Zones</td>
<td>55</td>
</tr>
<tr>
<td>Supported Regions &amp; Endpoints</td>
<td>56</td>
</tr>
<tr>
<td>Locating Your Nodes</td>
<td>59</td>
</tr>
<tr>
<td>Managing Maintenance</td>
<td>59</td>
</tr>
<tr>
<td>Managing Nodes</td>
<td>60</td>
</tr>
<tr>
<td>Redis Nodes and Shards</td>
<td>61</td>
</tr>
<tr>
<td>Connecting to Nodes</td>
<td>62</td>
</tr>
<tr>
<td>Reserved Nodes</td>
<td>64</td>
</tr>
<tr>
<td>Supported Node Types</td>
<td>65</td>
</tr>
<tr>
<td>Replacing Nodes</td>
<td>69</td>
</tr>
<tr>
<td>Managing Your ElastiCache Clusters</td>
<td>72</td>
</tr>
<tr>
<td>Supported Redis Versions</td>
<td>73</td>
</tr>
<tr>
<td>Other ElastiCache Cluster Operations</td>
<td>73</td>
</tr>
<tr>
<td>Creating a Cluster</td>
<td>74</td>
</tr>
<tr>
<td>Viewing a Cluster's Details</td>
<td>89</td>
</tr>
<tr>
<td>Modifying a Cluster</td>
<td>100</td>
</tr>
<tr>
<td>Rebooting a Cluster</td>
<td>103</td>
</tr>
<tr>
<td>Adding Nodes to a Cluster</td>
<td>104</td>
</tr>
<tr>
<td>Removing Nodes from a Cluster</td>
<td>110</td>
</tr>
<tr>
<td>Canceling Pending Add or Delete Node Operations</td>
<td>116</td>
</tr>
<tr>
<td>Deleting a Cluster</td>
<td>117</td>
</tr>
<tr>
<td>Accessing Your Cluster or Replication Group</td>
<td>119</td>
</tr>
<tr>
<td>Determine the Cluster's Platform</td>
<td>119</td>
</tr>
<tr>
<td>Grant Access to Your Cluster or Replication Group</td>
<td>121</td>
</tr>
<tr>
<td>Shards</td>
<td>127</td>
</tr>
<tr>
<td>Finding a Shard's ID</td>
<td>128</td>
</tr>
<tr>
<td>Replication Across AWS Regions Using Global Datastore</td>
<td>130</td>
</tr>
<tr>
<td>Overview</td>
<td>130</td>
</tr>
<tr>
<td>Prerequisites and Limitations</td>
<td>132</td>
</tr>
<tr>
<td>Using Global Datastores (Console)</td>
<td>132</td>
</tr>
<tr>
<td>Using Global Datastores (CLI)</td>
<td>140</td>
</tr>
<tr>
<td>High Availability Using Replication Groups</td>
<td>142</td>
</tr>
<tr>
<td>Understanding Redis Replication</td>
<td>144</td>
</tr>
<tr>
<td>Replication: Redis (Cluster Mode Disabled) vs. Redis (Cluster Mode Enabled)</td>
<td>145</td>
</tr>
<tr>
<td>Minimizing Downtime: Multi-AZ with Automatic Failover</td>
<td>148</td>
</tr>
<tr>
<td>How Synchronization and Backup are Implemented</td>
<td>161</td>
</tr>
<tr>
<td>Creating a Replication Group</td>
<td>162</td>
</tr>
<tr>
<td>Viewing a Replication Group's Details</td>
<td>182</td>
</tr>
<tr>
<td>Finding Replication Group Endpoints</td>
<td>187</td>
</tr>
<tr>
<td>Modifying a Replication Group</td>
<td>188</td>
</tr>
<tr>
<td>Deleting a Replication Group</td>
<td>190</td>
</tr>
<tr>
<td>Changing the Number of Replicas</td>
<td>191</td>
</tr>
<tr>
<td>Promoting a Read Replica</td>
<td>203</td>
</tr>
<tr>
<td>Finding Connection Endpoints</td>
<td>204</td>
</tr>
<tr>
<td>Finding Redis (Cluster Mode Disabled) Cluster Endpoints (Console)</td>
<td>206</td>
</tr>
<tr>
<td>Finding Endpoints for a Redis (Cluster Mode Enabled) Endpoints (Console)</td>
<td>208</td>
</tr>
<tr>
<td>Finding Endpoints (AWS CLI)</td>
<td>210</td>
</tr>
<tr>
<td>Finding Endpoints (ElastiCache API)</td>
<td>213</td>
</tr>
<tr>
<td>Backup and Restore</td>
<td>213</td>
</tr>
<tr>
<td>Constraints</td>
<td>214</td>
</tr>
<tr>
<td>Costs</td>
<td>215</td>
</tr>
<tr>
<td>Performance Impact of Backups</td>
<td>215</td>
</tr>
</tbody>
</table>
Managing Reserved Memory ................................................................. 466
How Much Reserved Memory Do You Need? ........................................ 466
Parameters to Manage Reserved Memory ............................................. 466
Specifying Your Reserved Memory Management Parameter .................. 469
Mitigating Out-of-Disk-Space Issues When Using Redis AOF ................. 471
Enabling Redis Multi-AZ as a Better Approach to Fault Tolerance .......... 471
Best Practices: Online Cluster Resizing .............................................. 472
Best Practices: Minimizing Downtime During Maintenance ................. 473
Self-Service Updates ........................................................................... 474
Managing the Service Updates ............................................................ 474
Applying the Self-Service Updates ...................................................... 474
Stopping the Self-Service Updates ...................................................... 480
Reference ........................................................................................... 483
Using the ElastiCache API ................................................................. 483
Using the Query API .......................................................................... 485
Available Libraries ........................................................................... 486
Troubleshooting Applications ............................................................. 486
Logging Amazon ElastiCache API Calls with AWS CloudTrail ............. 486
Set Up the AWS CLI for ElastiCache .................................................. 490
Prerequisites ..................................................................................... 490
Getting the Command Line Tools ....................................................... 491
Setting Up the Tools ......................................................................... 491
Providing Credentials for the Tools .................................................... 492
Environmental Variables .................................................................... 493
Error Messages ................................................................................ 493
Notifications ..................................................................................... 494
General ElastiCache Notifications ...................................................... 494
ElastiCache for Redis Specific Notifications ....................................... 495
ElastiCache for Redis Documentation History ..................................... 496
AWS Glossary ................................................................................... 509
What Is Amazon ElastiCache for Redis?

Welcome to the Amazon ElastiCache for Redis 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. 20) in this guide.

Existing applications that use Redis 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.

ElastiCache for Redis has multiple features that help make the service more reliable for critical production deployments:

- Automatic detection of and recovery from cache node failures.
- Multi-AZ with automatic failover for a failed primary cluster to a read replica, in Redis clusters that support replication.
- Redis (cluster mode enabled) supports partitioning your data across up to 90 shards.
- For Redis version 3.2 and later, all versions support encryption in transit and encryption at rest encryption with authentication. This support helps you build HIPAA-compliant applications.
- Flexible Availability Zone placement of nodes and clusters for increased fault tolerance.
- Integration with other AWS services such as Amazon EC2, Amazon CloudWatch, AWS CloudTrail, and Amazon SNS. This integration helps provide a managed in-memory caching solution that is high-performance and highly secure.

Topics
- Common ElastiCache Use Cases and How ElastiCache Can Help (p. 2)
- Amazon ElastiCache Resources (p. 8)
- ElastiCache for Redis Components and Features (p. 10)
- ElastiCache for Redis Terminology (p. 17)
- Tools for Managing Your Implementation (p. 19)
Whether serving the latest news, a top-10 leaderboard, 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.

Topics
- In-Memory Data Store (p. 2)
- Gaming Leaderboards (Redis Sorted Sets) (p. 4)
- Messaging (Redis Pub/Sub) (p. 5)
- Recommendation Data (Redis Hashes) (p. 7)
- Other Redis Uses (p. 7)
- ElastiCache Customer Testimonials (p. 7)

**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 Redis User Guide*
Gaming Leaderboards (Redis Sorted Sets)

Redis sorted sets move the computational complexity of leaderboards from your application to your Redis cluster.

Leaderboards, such as the top 10 scores for a game, are computationally complex. This is especially true when there is a large number of concurrent players and continually changing scores. Redis sorted sets guarantee both uniqueness and element ordering. Using Redis sorted sets, each time a new element is added to the sorted set it’s reranked in real time. It’s then added to the set in its correct numeric order.

In the following diagram, you can see how an ElastiCache for Redis gaming leaderboard works.

**Example - Redis Leaderboard**

In this example, four gamers and their scores are entered into a sorted list using ZADD. The command ZREVRANGEBYSCORE lists the players by their score, high to low. Next, ZADD is used to update June’s score by overwriting the existing entry. Finally, ZREVRANGEBYSCORE lists the players by their score, high to low. The list shows that June has moved up in the rankings.

```
ZADD leaderboard 132 Robert
ZADD leaderboard 231 Sandra
ZADD leaderboard 32 June
ZADD leaderboard 381 Adam
ZREVRANGEBYSCORE leaderboard +inf -inf
1) Adam
2) Sandra
3) Robert
4) June

ZADD leaderboard 232 June
ZREVRANGEBYSCORE leaderboard +inf -inf
1) Adam
2) June
3) Sandra
4) Robert
```

The following command tells June where she ranks among all the players. Because ranking is zero-based, ZREVRANK returns a 1 for June, who is in second position.
Messaging (Redis Pub/Sub)

When you send an email message, you send it to one or more specified recipients. In the pub/sub paradigm, you send a message to a specific channel not knowing who, if anyone, receives it. The people who get the message are those who are subscribed to the channel. For example, suppose that you subscribe to the `news.sports.golf` channel. You and all others subscribed to the `news.sports.golf` channel get any messages published to `news.sports.golf`.

Redis pub/sub functionality has no relation to any key space. Therefore, it doesn't interfere on any level. In the following diagram, you can find an illustration of ElastiCache for Redis messaging.

Subscribing

To receive messages on a channel, you subscribe to the channel. You can subscribe to a single channel, multiple specified channels, or all channels that match a pattern. To cancel a subscription, you unsubscribe from the channel specified when you subscribed to it. Or, if you subscribed using pattern matching, you unsubscribe using the same pattern that you used before.

Example - Subscription to a Single Channel

To subscribe to a single channel, use the SUBSCRIBE command specifying the channel you want to subscribe to. In the following example, a client subscribes to the `news.sports.golf` channel.

```
SUBSCRIBE news.sports.golf
```

After a while, the client cancels their subscription to the channel using the UNSUBSCRIBE command specifying the channel to unsubscribe from.

```
UNSUBSCRIBE news.sports.golf
```
Example - Subscriptions to Multiple Specified Channels

To subscribe to multiple specific channels, list the channels with the SUBSCRIBE command. In the following example, a client subscribes to the `news.sports.golf`, `news.sports.soccer`, and `news.sports.skiing` channels.

```
SUBSCRIBE news.sports.golf news.sports.soccer news.sports.skiing
```

To cancel a subscription to a specific channel, use the UNSUBSCRIBE command and specify the channel to unsubscribe from.

```
UNSUBSCRIBE news.sports.golf
```

To cancel subscriptions to multiple channels, use the UNSUBSCRIBE command and specify the channels to unsubscribe from.

```
UNSUBSCRIBE news.sports.golf news.sports.soccer
```

To cancel all subscriptions, use UNSUBSCRIBE and specify each channel. Or use UNSUBSCRIBE and don't specify a channel.

```
UNSUBSCRIBE news.sports.golf news.sports.soccer news.sports.skiing
```

or

```
UNSUBSCRIBE
```

Example - Subscriptions Using Pattern Matching

Clients can subscribe to all channels that match a pattern by using the PSUBSCRIBE command.

In the following example, a client subscribes to all sports channels. You don't list all the sports channels individually, as you do using SUBSCRIBE. Instead, with the PSUBSCRIBE command you use pattern matching.

```
PSUBSCRIBE news.sports.*
```

Example Canceling Subscriptions

To cancel subscriptions to these channels, use the PUNSUBSCRIBE command.

```
PUNSUBSCRIBE news.sports.*
```

**Important**
The channel string sent to a [P]SUBSCRIBE command and to the [P]UNSUBSCRIBE command must match. You can't PSUBSCRIBE to `news.*` and PUNSUBSCRIBE from `news.sports.*` or UNSUBSCRIBE from `news.sports.golf`.

Publishing

To send a message to all subscribers to a channel, use the PUBLISH command, specifying the channel and the message. The following example publishes the message, "It's Saturday and sunny. I'm headed to the links." to the `news.sports.golf` channel.
A client can't publish to a channel that it's subscribed to.

For more information, see Pub/Sub in the Redis documentation.

Recommendation Data (Redis Hashes)

Using INCR or DECR in Redis makes compiling recommendations simple. Each time a user "likes" a product, you increment an `item:productID:like` counter. Each time a user "dislikes" a product, you increment an `item:productID:dislike` counter. Using Redis hashes, you can also maintain a list of everyone who has liked or disliked a product. The following diagram illustrates an ElastiCache for Redis real-time analytics store.

Example - Likes and Dislikes

```
INCR item:38923:likes
HSET item:38923:ratings Susan 1
INCR item:38923:dislikes
HSET item:38923:ratings Tommy -1
```

Other Redis Uses

The blog post How to take advantage of Redis just adding it to your stack by Salvatore Sanfilippo discusses a number of common database concerns and how they can be easily solved using Redis. This approach removes load from your database and improves performance.

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. 35) for additional ElastiCache customer use cases.
Amazon ElastiCache 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. 35) section has videos that introduce you to Amazon ElastiCache. The videos cover common use cases for ElastiCache and demo how to use ElastiCache to reduce latency and improve throughput for your applications.
- **Getting started** – The Getting Started with Amazon ElastiCache for Redis (p. 24) section includes an example that walks you through the process of creating a cache cluster. The example includes how to authorize access to the cache cluster, connect to a cache node, and delete the cache cluster.
- **Performance at scale** – The Performance at Scale with Amazon ElastiCache whitepaper addresses caching strategies that help your application to perform well at scale.

After you complete the preceding sections, read these sections:

- **Choosing Your Node Size** (p. 76)
  
  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. 458)
  
  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. 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. 483)
  
  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 Redis Components and Features

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

Topics
- ElastiCache Nodes (p. 10)
- ElastiCache for Redis Shards (p. 10)
- ElastiCache for Redis Clusters (p. 11)
- ElastiCache for Redis Replication (p. 12)
- AWS Regions and Availability Zones (p. 13)
- ElastiCache for Redis Endpoints (p. 13)
- ElastiCache Parameter Groups (p. 14)
- ElastiCache for Redis Security (p. 14)
- ElastiCache Security Groups (p. 14)
- ElastiCache Subnet Groups (p. 15)
- ElastiCache for Redis Backups (p. 15)
- ElastiCache Events (p. 15)

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 the engine and version that was chosen when you created your cluster. 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 Redis Clusters (p. 250).

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. 65).

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 much-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. You can then purchase pay-as-you-go nodes for the times you occasionally need to add nodes. For more information on reserved nodes, see ElastiCache ReservedNodes (p. 64).

For more information on nodes, see Managing Nodes (p. 60).

ElastiCache for Redis Shards

A Redis shard (called a node group in the API and CLI) is a grouping of one to six related nodes. A Redis (cluster mode disabled) cluster always has one shard. A Redis (cluster mode enabled) cluster can have 1–90 shards.

A multiple node shard implements replication by have one read/write primary node and 1–5 replica nodes. For more information, see High Availability Using Replication Groups (p. 142).

For more information on shards, see Working with Shards (p. 127).
ElastiCache for Redis Clusters

A Redis cluster is a logical grouping of one or more ElastiCache for Redis Shards (p. 10). Data is partitioned across the shards in a Redis (cluster mode enabled) cluster.

Many ElastiCache operations are targeted at clusters:

- Creating a cluster
- Modifying a cluster
- Taking snapshots of a cluster (all versions of Redis)
- 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. 72) and Managing Nodes (p. 60)
  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. To exceed certain of these limits, you can make a request using the Amazon ElastiCache Cache Node request form.
- Mitigating Failures (p. 454)
  Information about improving the fault tolerance of your clusters and replication groups.

Typical Cluster Configurations

A Redis cluster contains 1–90 shards (in the API, called node groups), each of which is a partition of your data. Redis (cluster mode disabled) always has just one shard.

Following are typical cluster configurations.

Redis Clusters

A Redis (cluster mode enabled) cluster contains 1–90 shards (in the API and CLI, called node groups). Redis (cluster mode disabled) clusters always contain just one shard (in the API and CLI, one node group). A Redis shard contains one to six nodes. If there is more than one node in a shard, the shard supports replication. In this case, one node is the read/write primary node and the others are read-only replica nodes.

For improved fault tolerance, we recommend having at least two nodes in a Redis cluster and enabling Multi-AZ with automatic failover. For more information, see Mitigating Failures (p. 454).

As demand upon your Redis (cluster mode disabled) cluster changes, you can scale up or down. To do this, you move your cluster to a different node instance type. If your application is read intensive, we recommend adding read-only replicas Redis (cluster mode disabled) cluster. By doing this, you can spread the reads across a more appropriate number of nodes.

ElastiCache supports changing a Redis (cluster mode disabled) cluster's node type to a larger node type dynamically. For information on scaling up or down, see Scaling Single-Node Clusters for Redis (Cluster Mode Disabled) (p. 252) or Scaling Redis (Cluster Mode Disabled) Clusters with Replica Nodes (p. 264).
ElastiCache for Redis Replication

Before you continue reading here, see ElastiCache for Redis Terminology (p. 17) to better understand the differences in terminology between the ElastiCache console and the ElastiCache API and AWS CLI.

Replication is implemented by grouping from two to six nodes in a shard (in the API and CLI, called a node group). One of these nodes is the read/write primary node. All the other nodes are read-only replica nodes.

Each replica node maintains a copy of the data from the primary node. Replica nodes use asynchronous replication mechanisms to keep synchronized with the primary node. Applications can read from any node in the cluster but can write only to primary nodes. Read replicas enhance scalability by spreading reads across multiple endpoints. Read replicas also improve fault tolerance by maintaining multiple copies of the data. Locating read replicas in multiple Availability Zones further improves fault tolerance. For more information on fault tolerance, see Mitigating Failures (p. 454).

Redis (cluster mode disabled) clusters support one shard (in the API and CLI, called a node group). Redis (cluster mode enabled) clusters support 1–90 shards (in the API and CLI, called node groups).

Replication from the API and CLI perspective uses different terminology to maintain compatibility with previous versions, but the results are the same. The following table shows the API and CLI terms for implementing replication.

Comparing Replication: Redis (cluster mode disabled) and Redis (cluster mode enabled)

In the following table, you can find a comparison of the features of Redis (cluster mode disabled) and Redis (cluster mode enabled) replication groups.

<table>
<thead>
<tr>
<th></th>
<th>Redis (cluster mode disabled)</th>
<th>Redis (cluster mode enabled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shards (node groups)</td>
<td>1</td>
<td>1–90</td>
</tr>
<tr>
<td>Replicas for each shard (node group)</td>
<td>0–5</td>
<td>0–5</td>
</tr>
<tr>
<td>Data partitioning</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Add/Delete replicas</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Add/Delete node groups</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Supports scale up</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Supports engine upgrades</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Promote replica to primary</td>
<td>Yes</td>
<td>Automatic</td>
</tr>
<tr>
<td>Multi-AZ with automatic failover</td>
<td>Optional</td>
<td>Required</td>
</tr>
<tr>
<td>Backup/Restore</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes:

- If any primary has no replicas and the primary fails, you lose all that primary's data.
- You can use backup and restore to migrate to Redis (cluster mode enabled).
- You can use backup and restore to resize your Redis (cluster mode enabled) cluster.
All of the shards (in the API and CLI, node groups) and nodes must reside in the same AWS Region. However, you can provision the individual nodes in multiple Availability Zones within that AWS Region.

Read replicas guard against potential data loss because your data is replicated over two or more nodes—the primary and one or more read replicas. For greater reliability and faster recovery, we recommend that you create one or more read replicas in different Availability Zones. In addition, enable Multi-AZ with automatic failover instead of using AOF. AOF is disabled when Multi-AZ with automatic failover is enabled. For more information, see Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148).

Replication: Limits and Exclusions
- AOF is not supported on node type cache.t1.micro and cache.t2. For nodes of these types, the appendonly parameter value is ignored.
- Multi-AZ with automatic failover is not supported on node types T1.

For more information on AOF and Multi-AZ, see Mitigating Failures (p. 454).

AWS Regions and Availability Zones

Amazon ElastiCache 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. You can use these 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. 55). In the following diagram, you can see a high-level view of how AWS Regions and Availability Zones work.

For information on AWS Regions supported by ElastiCache and their endpoints, see Supported Regions & Endpoints (p. 56).

ElastiCache for Redis Endpoints

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

Single Node Endpoints for Redis (Cluster Mode Disabled)

The endpoint for a single node Redis cluster is used to connect to the cluster for both reads and writes.
Multi-Node Endpoints for Redis (Cluster Mode Disabled)

A multiple node Redis (cluster mode disabled) cluster has two types of endpoints. The primary endpoint always connects to the primary node in the cluster, even if the specific node in the primary role changes. Use the primary endpoint for all writes to the cluster.

The read endpoint in a Redis (cluster mode disabled) cluster always points to a specific node. Whenever you add or remove a read replica, you must update the associated node endpoint in your application.

Redis (Cluster Mode Enabled) Endpoints

A Redis (cluster mode enabled) cluster has a single configuration endpoint. By connecting to the configuration endpoint, your application is able to discover the primary and read endpoints for each shard in the cluster.

For more information, see Finding Connection Endpoints (p. 204).

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 (CLI: describe-engine-default-parameters).

For more detailed information on ElastiCache parameter groups, see Configuring Engine Parameters Using Parameter Groups (p. 298).

ElastiCache for Redis Security

For enhanced security, ElastiCache for Redis node access is restricted to applications running on the Amazon EC2 instances that you allow. You can control the Amazon EC2 instances that can access your cluster by using subnet groups or security groups.

By default, all new ElastiCache for Redis 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. These enable you to authorize Amazon EC2 instances running within specific Amazon EC2 security groups.

In addition to restricting node access, ElastiCache for Redis supports TLS and in-place encryption for nodes running specified versions of ElastiCache for Redis. For more information, see the following:

- Data Security in Amazon ElastiCache (p. 340)
- HIPAA Eligibility (p. 452)
- Authenticating Users with the Redis AUTH Command (p. 354)

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 run your ElastiCache nodes in a virtual private cloud (VPC) based on Amazon VPC, you control access to your cache clusters with Amazon VPC security groups. These are different from ElastiCache security groups. For
more information on using ElastiCache with Amazon VPC, see Amazon VPCs and ElastiCache Security (p. 358).

With ElastiCache, you can 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, 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. You can also do this by using the ElastiCache Management Console, the ModifyCacheCluster API operation, or the modify-cache-cluster AWS CLI command.

Important
Access control based on IP ranges is currently not enabled for clusters. All clients to a cluster must be within the Amazon EC2 network, and authorized by using security groups as described previously.

For more information about security groups, see Security Groups: EC2-Classic (p. 382).

ElastiCache Subnet Groups

A subnet group is a collection of subnets (typically private) that you can designate for your clusters running in an 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 the following:

- Amazon VPCs and ElastiCache Security (p. 358)
- Step 2: Authorize Access (p. 28)
- Subnets and Subnet Groups (p. 374)

ElastiCache for Redis Backups

A backup is a point-in-time copy of a Redis cluster. Backups can be used to restore an existing cluster or to seed a new cluster. Backups consist of all the data in a cluster plus some metadata.

Depending upon the version of Redis running on your cluster, the backup process requires differing amounts of reserved memory to succeed. For more information, see the following:

- Backup and Restore for ElastiCache for Redis (p. 213)
- How Synchronization and Backup are Implemented (p. 161)
- Performance Impact of Backups (p. 215)
- Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464)

ElastiCache Events

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

For more information on ElastiCache events, see Monitoring ElastiCache Events (p. 423).
ElastiCache for Redis Terminology

In October 2016, Amazon ElastiCache launched support for Redis 3.2. At that point, we added support for partitioning your data across up to 90 shards (called node groups in the ElastiCache API and AWS CLI). To preserve compatibility with previous versions, we extended API version 2015-02-02 operations to include the new Redis functionality.

At the same time, we began using terminology in the ElastiCache console that is used in this new functionality and common across the industry. These changes mean that at some points, the terminology used in the API and CLI might be different from the terminology used in the console. The following list identifies terms that might differ between the API and CLI and the console.

**Cache cluster or node vs. node**

There is a one-to-one relationship between a node and a cache cluster when there are no replica nodes. Thus, the ElastiCache console often used the terms interchangeably. The console now uses the term *node* throughout. The one exception is the Create Cluster button, which launches the process to create a cluster with or without replica nodes.

The ElastiCache API and AWS CLI continue to use the terms as they have in the past.

**Cluster vs. replication group**

The console now uses the term *cluster* for all ElastiCache for Redis clusters. The console uses the term *cluster* in all these circumstances:
- When the cluster is a single node Redis cluster.
- When the cluster is a Redis (cluster mode disabled) cluster that supports replication within a single shard (in the API and CLI, called a *node group*).
- When the cluster is a Redis (cluster mode enabled) cluster that supports replication within 1–90 shards.

The following diagram illustrates the various topologies of ElastiCache for Redis clusters from the console's perspective.

**ElastiCache for Redis: Console View**

The ElastiCache API and AWS CLI operations still distinguish single node ElastiCache for Redis clusters from multi-node replication groups. The following diagram illustrates the various ElastiCache for Redis topologies from the ElastiCache API and AWS CLI perspective.
Tools for Managing Your Implementation

When 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. 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. These show cache engine activity, memory and CPU utilization, and 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. You also write code to parse results from ElastiCache and handle any errors. For more information about the API, see Using the ElastiCache API (p. 483).

Additional Resources

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

- Managing Your ElastiCache for Redis Implementation (p. 44)
- Internetwork Traffic Privacy (p. 357)
- Logging and Monitoring in Elasticache (p. 411)
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>Data partitioning</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Cluster is modifiable</td>
<td>Yes</td>
<td>No</td>
<td>Yes - 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 Redis

Following, you can find topics that lead you through creating, granting access to, connecting to, and finally deleting a Redis (cluster mode disabled) cluster using the ElastiCache Management Console. As part of this, the section starts by helping you determine the requirements for your cluster and create your own AWS account.

Amazon ElastiCache supports high availability through the use of Redis replication groups. For information about Redis replication groups and how to create them, see High Availability Using Replication Groups (p. 142).

Beginning with Redis version 3.2, ElastiCache Redis supports partitioning your data across multiple node groups, with each node group implementing a replication group. This exercise creates a standalone Redis cluster.

Topics
- Determine Requirements (p. 24)
- Setting Up (p. 24)
- Step 1: Launch a Cluster (p. 26)
- Step 2: Authorize Access (p. 28)
- Step 3: Connect to a Cluster’s Node (p. 29)
- Step 4: Delete Your Cluster (Avoid Unnecessary Charges) (p. 32)
- Where Do I Go From Here? (p. 33)

Determine Requirements

Before you create a Redis cluster or replication group, you should always determine the requirements for the cluster or replication group so that when you create it, 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. 74).

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. 24)
- Set Up Your Permissions (New ElastiCache Users Only) (p. 25)

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. 398)
- Using Service-Linked Roles for Amazon ElastiCache (p. 401)
Step 1: Launch a Cluster

The cluster you're about to launch will be live, and not running in a sandbox. You will incur the standard ElastiCache usage fees for the instance until you delete it. The total charges will be 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. 375).

**To create a standalone Redis (cluster mode disabled) cluster**

1. Sign in to the AWS Management Console and open the Amazon ElastiCache console at https://console.aws.amazon.com/elasticache/.
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 that you want to launch this cluster in.
4. For **Cluster engine**, choose Redis.
5. Make sure that **Cluster Mode enabled (Scale Out)** is not chosen.
6. Complete the **Redis settings** section as follows:
   
   a. For **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 Redis 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, 6379. If you have a reason to use a different port, enter 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. 300).

   e. For **Node type**, choose the node type that you want to use for this cluster. For this exercise, above the table choose the t2 instance family, choose **cache.t2.small**, and finally choose **Save**.

   For more information, see Choosing Your Node Size (p. 76).

   f. From **Number of replicas**, choose the number of read replicas you want for this cluster. Because in this exercise we're creating a standalone cluster, choose **None**.

   When you choose **None**, the **Replication group description** field disappears.

7. Choose **Advanced Redis settings** and complete the section as follows:
Note
The Advanced Redis settings details are slightly different if you are creating a Redis (cluster mode enabled) replication group. For a step-by-step walkthrough to create a Redis (cluster mode enabled) replication group, see Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (p. 175).

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

For more information, see Subnets and Subnet Groups (p. 374).

b. For Availability zone(s), you have two options.
   - No preference – ElastiCache chooses the Availability Zone.
   - Specify availability zones – You specify the Availability Zone for your cluster.

For this exercise, choose Specify availability zones and then choose an Availability Zone from the list below Primary.

For more information, see Choosing Regions and Availability Zones (p. 55).

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

For more information, see Amazon VPCs and ElastiCache Security (p. 358).

d. If you are going to seed your cluster with data from a .RDB file, in the Seed RDB file S3 location box, enter the Amazon S3 location of the .RDB file.

For more information, see Seeding a New Cluster with an Externally Created Backup (p. 242).

e. Because this is not a production cluster, clear the Enable automatic backups check box.

For more information on Redis backup and restore, see Backup and Restore for ElastiCache for Redis (p. 213).

f. The Maintenance window is the time, generally an hour, each week where ElastiCache schedules system maintenance on your cluster. You can allow ElastiCache to specify the day and time for your maintenance window (No preference), or you can specify the day and time yourself (Specify maintenance window). If you choose Specify maintenance window, specify the Start day, Start time, and Duration (in hours) for your maintenance window. For this exercise, choose No preference.

For more information, see Managing Maintenance (p. 59).

g. For Notifications, leave it as Disabled.

8. Choose Create cluster 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 AWS Region, or even your corporate network, see the following:

- Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 363)
- Accessing ElastiCache Resources from Outside AWS (p. 123)

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 or Replication Group (p. 119).
Step 3: Connect to a Cluster's Node

Before you continue, complete Step 2: Authorize Access (p. 28).

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. 28).

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. 28)), 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 for the engine and cluster type you're running. When you find the endpoint you need, copy it to your clipboard for use in Step 3.2.

- Finding Connection Endpoints (p. 204)
- Finding a Redis (Cluster Mode Disabled) Cluster's Endpoints (Console) (p. 206)—You need the primary endpoint of a replication group or the node endpoint of a standalone node.
- Finding Endpoints for a Redis (Cluster Mode Enabled) Cluster (Console) (p. 208)—You need the cluster's Configuration endpoint.
- Finding Endpoints (AWS CLI) (p. 210)
- Finding Endpoints (ElastiCache API) (p. 213)

Step 3.2: Connect to a Redis Cluster or Replication Group (Linux)

Now that you have the endpoint you need, you can log in to an EC2 instance and connect to the cluster or replication group.

In the following example, you use the redis-cli utility to connect to a cluster that is not encryption enabled and running Redis. For more information about Redis and available Redis commands, see Redis commands webpage.

To connect to a Redis cluster that is not encryption-enabled using redis-cli

1. Connect to your Amazon EC2 instance using the connection utility of your choice. For instructions on how to connect to an Amazon EC2 instance, see the Amazon EC2 Getting Started Guide.
2. Download and install the GNU Compiler Collection (gcc).

   At the command prompt of your EC2 instance, type the following command then, at the confirmation prompt, type y.

   ```bash
   sudo yum install gcc
   ```

   Doing this produces output similar to the following.

   ```bash
   Loaded plugins: priorities, security, update-motd, upgrade-helper
   ```
Step 3.2: Connect to a Redis Cluster or Replication Group (Linux)

Setting up Install Process
Resolving Dependencies
--> Running transaction check

...(output omitted)...

Total download size: 27 M
Installed size: 53 M
Is this ok [y/N]: y

Downloading Packages:

(1/11): binutils-2.22.52.0.1-10.36.amzn1.x86_64.rpm  | 5.2 MB  00:00
(2/11): cpp46-4.6.3-2.67.amzn1.x86_64.rpm           | 4.8 MB  00:00
(3/11): gcc-4.6.3-3.10.amzn1.noarch.rpm             | 2.8 kB  00:00

...(output omitted)...

Complete!

3. Download and compile the *redis-cli* utility. This utility is included in the Redis software distribution.

At the command prompt of your EC2 instance, type the following commands:

```
wget http://download.redis.io/redis-stable.tar.gz
```

```
tar xvzf redis-stable.tar.gz
```

```
cd redis-stable
```

```
made distclean  // Ubuntu systems only
```

```
made
```

4. At the command prompt of your EC2 instance, type the following command, substituting the endpoint of your cluster and port for what is shown in this example.

```
src/redis-cli -c -h mycachecluster.eaogs8.0001.usw2.cache.amazonaws.com -p 6379
```

This results in a Redis command prompt similar to the following.

```
redis mycachecluster.eaogs8.0001.usw2.cache.amazonaws.com 6379>
```

5. Run Redis commands.

You are now connected to the cluster and can run Redis commands like the following.

```
set a "hello"       // Set key "a" with a string value and no expiration
OK
get a                // Get value for key "a"
"hello"
get b                // Get value for key "b" results in miss
(nil)
set b "Good-bye" EX 5 "Good-bye"  // Set key "b" with a string value and a 5 second expiration
get b                // Get value for key "b"
"Good-bye"
get b                // wait >= 5 seconds
(nil)                // key has expired, nothing returned
quit                 // Exit from redis-cli
```
Step 3.2: Connect to a Redis Cluster or Replication Group (Windows)

In order to connect to the Redis Cluster from an EC2 Windows instance using the Redis CLI, you must download the `redis-cli` package and use `redis-cli.exe` to connect to the Redis Cluster from an EC2 Windows instance.

In the following example, you use the `redis-cli` utility to connect to a cluster that is not encryption enabled and running Redis. For more information about Redis and available Redis commands, see Redis commands webpage.

To connect to a Redis cluster that is not encryption-enabled using `redis-cli`

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

2. Copy and paste the link https://github.com/microsoftarchive/redis/releases/download/win-3.0.504/Redis-x64-3.0.504.zip in an Internet browser to download the zip file for the Redis client from the available release at Github https://github.com/microsoftarchive/redis/releases/tag/win-3.0.504

   Extract the zip file to your desired folder/path.

   Open the Command Prompt and change to the Redis directory and run the command

   ```
c:\Redis\redis-cli -h Redis_Cluster_Endpoint -p 6379.
   ```

   For example:

   ```
c:\Redis\redis-cli -h cmd.xxxxxxx.ng.0001.usw2.cache.amazonaws.com -p 6379
   ```

3. Run Redis commands.

   You are now connected to the cluster and can run Redis commands like the following.

   ```
   set a "hello" // Set key "a" with a string value and no expiration
   OK
   get a "hello" // Get value for key "a"
   get b (nil) // Get value for key "b" results in miss
   set b "Good-bye" EX 5 "Good-bye" // Set key "b" with a string value and a 5 second expiration
   get b "Good-bye" // Get value for key "b"
   // wait >= 5 seconds
   get b (nil) // key has expired, nothing returned
   quit // Exit from redis-cli
   ```
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, complete at least as far as Step 1: Launch a Cluster (p. 26).

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. To see a list of all your clusters running Redis, in the navigation pane, choose Redis.
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 button. To delete multiple clusters, repeat this process for each cluster. You do not need to wait for one cluster to finish deleting before you delete another 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 an ElastiCache for Redis 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 Node Size (p. 76)

  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. 458)

  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. 35)
- 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. 35)
- Advanced Videos (p. 35)

Introductory Videos

The following videos introduce you to Amazon ElastiCache.

Topics
- DAT204—Building Scalable Applications on AWS NoSQL Services (re:Invent 2015) (p. 35)
- DAT207—Accelerating Application Performance with Amazon ElastiCache (AWS re:Invent 2013) (p. 35)

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
- DAT305—Amazon ElastiCache Deep Dive (re:Invent 2017) (p. 36)
- DAT306—Amazon ElastiCache Deep Dive (re:Invent 2016) (p. 36)
- DAT317—How IFTTT uses ElastiCache for Redis to Predict Events (re:Invent 2016) (p. 36)
• DAT407—Amazon ElastiCache Deep Dive (re:Invent 2015) (p. 36)
• SDD402—Amazon ElastiCache Deep Dive (re:Invent 2014) (p. 36)
• DAT307—Deep Dive into Amazon ElastiCache Architecture and Design Patterns (re:Invent 2013) (p. 37)

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)

DAT317—How IFTTT uses ElastiCache for Redis to Predict Events (re:Invent 2016)

IFTTT is a free service that empowers people to do more with the services they love, from automating simple tasks to transforming how someone interacts with and controls their home. IFTTT uses ElastiCache for Redis to store transaction run history and schedule predictions as well as indexes for log documents on Amazon S3. View this session to learn how the scripting power of Lua and the data types of Redis allowed people to accomplish something they wouldn't have been able to elsewhere.

DAT317—How IFTTT uses ElastiCache for Redis to Predict Events (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.

DAT402—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.

Online Migration to ElastiCache

By using Online Migration, you can migrate your data from self-hosted Redis on Amazon EC2 to Amazon ElastiCache.

Overview

To migrate your data from Redis running on Amazon EC2 to Amazon ElastiCache requires an existing or newly created Amazon ElastiCache deployment. The deployment must have a configuration that is ready for migration. It also should be in line with the configuration that you want, including attributes like instance type and number of replicas.

Important
We strongly recommend you read the following sections in their entirety before beginning the online migration process.

The migration begins when you call the StartMigration API operation or AWS CLI command. The migration process makes the master node of the ElastiCache for Redis cluster a replica to your source Redis cluster on EC2. Using Redis replication, data is synced between your source Redis and ElastiCache. After the data is in sync, you are nearly ready to cut over to ElastiCache. At this point, you make changes on the application side so your application can call ElastiCache post-migration.

After the client-side changes are ready, call the CompleteMigration API operation. This API operation promotes your ElastiCache deployment to your primary Redis deployment with primary and replica nodes (as applicable). Now you can redirect your client application to start writing data to ElastiCache. Throughout the migration, you can check the status of replication by running the redis-cli INFO command on your Redis on EC2 nodes and on the ElastiCache primary node.

Migration Steps

The following topics outline the process for migrating your data:

- Preparing Your Source and Target Redis Nodes for Migration (p. 38)
- Starting Migration (p. 39)
- Verifying the Data Migration Progress (p. 40)
- Completing the Data Migration (p. 41)

Preparing Your Source and Target Redis Nodes for Migration

You must ensure that all four of the pre-requisites mentioned below are satisfied before you start the migration from ElastiCache console, API or AWS CLI.
For more information, see Getting Started with Migrating Data Online from Redis on Amazon EC2 to Fully-Managed Amazon ElastiCache at 17:08 of the video.

To prepare your source and target Redis Nodes for migration

1. Identify the target ElastiCache deployment and make sure that you can migrate data to it.

   An existing or newly created ElastiCache deployment should meet the following requirements for migration:
   - It's cluster-mode disabled using Redis engine version 5.0.5 or higher.
   - It doesn't have either encryption in-transit or encryption at-rest enabled.
   - It has Multi-AZ with Auto-Failover enabled.
   - It has sufficient memory available to fit the data from your Redis on EC2 instance. To configure the right reserved memory settings, see Managing Reserved Memory (p. 466).
   - You can migrate directly from Redis versions 2.8.21 onward to Redis versions 5.0.5 onward.

2. Make sure that the configurations of your Redis on EC2 and the ElastiCache for Redis deployment are compatible.

   At a minimum, all the following in the target ElastiCache deployment should be compatible with your Redis configuration for Redis replication:
   - Your Redis cluster should be in cluster-mode disabled configuration.
   - You Redis on EC2 instance should not have Redis AUTH enabled.
   - Redis config protected-mode should be set to no.
   - If you have bind configuration in your Redis config, then it should be updated to allow requests from ElastiCache nodes.
   - The number of logical databases should be the same on the ElastiCache node and your Redis on EC2 instance. This value is set using databases in the Redis config.
   - Redis commands that perform data modification should not be renamed to allow replication of the data to succeed.
   - To replicate the data from your Redis cluster to ElastiCache, make sure that there is sufficient CPU and memory to handle this additional load. This load comes from the RDB file created by your Redis cluster and transferred over the network to ElastiCache node.

3. Make sure that your EC2 instance can connect with ElastiCache by doing the following:

   - Ensure that your EC2 instance's IP address is private.
   - Assign or create the ElastiCache deployment in the same VPC as your Redis on EC2 instance (recommended).
   - If the VPCs are different, set up VPC peering to allow access between the nodes. For more information on VPC peering, see Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 363).
   - The security group attached to your Redis on EC2 instance should allow inbound traffic from ElastiCache nodes.

4. Make sure that your application can direct traffic to ElastiCache nodes after migration of data is complete. For more information, see Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 363).

Starting Migration

After all prerequisites are complete, you can begin data migration using the AWS Management Console, ElastiCache API, or AWS CLI. The following example shows using the CLI.
Start migration by calling the start-migration command with the following parameters:

- `--replication-group-id` – Identifier of the target ElastiCache replication group
- `--customer-node-endpoint-list` – A list of endpoints with either DNS or IP addresses and the port where your source Redis on EC2 cluster is running. Because ElastiCache currently only supports cluster-mode disabled configuration, this list should contain one entry. If you have enabled chained replication, the endpoint can point to a replica instead of the master node in your Redis cluster.

The following is an example using the CLI.

```bash
aws elasticache start-migration --replication-group-id test-cluster --customer-node-endpoint-list "Address='10.0.0.241',Port=6379"
```

As you run this command, the ElastiCache primary node configures itself to become a replica of your Redis on EC2 instance. The status of ElastiCache cluster changes to `migrating` and data starts migrating from your Redis on EC2 instance to the ElastiCache primary node. Depending on the size of the data and load on your Redis instance, the migration can take a while to complete. You can check the progress of the migration by running the `redis-cli INFO` command on your Redis on EC2 instance and ElastiCache primary node.

After successful replication, all writes to your Redis on EC2 instance propagate to the ElastiCache cluster. You can use ElastiCache nodes for reads. However, you can't write to the ElastiCache cluster. If the ElastiCache primary node has other replica nodes connected to it, these replica nodes continue to replicate from the ElastiCache primary node. This way, all the data from your Redis on EC2 instance gets replicated to all the nodes in ElastiCache cluster.

If the ElastiCache primary node can’t become a replica of your Redis on EC2 instance, it retries several times before eventually promoting itself back to master. The status of ElastiCache cluster then changes to `available`, and a replication group event about the failure to initiate the migration is sent. To troubleshoot such a failure, check the following:

- Look at the replication group event. Use any specific information from the event to fix the migration failure.
- If the event doesn’t provide any specific information, make sure that you have followed the guidelines in Preparing Your Source and Target Redis Nodes for Migration (p. 38).
- Ensure that the routing configuration for your VPC and subnets allows traffic between ElastiCache nodes and your Redis on EC2 instance.
- Ensure the security group attached to your Redis on EC2 instance allows input bound traffic from ElastiCache nodes.
- Check Redis logs for your Redis on EC2 instance for more information about failures specific to replication.

**Verifying the Data Migration Progress**

After the migration has begun, you can do the following to track its progress:

- Verify that Redis `master_link_status` is up in the `INFO` command on ElastiCache primary node. You can also find this information in the ElastiCache console. Select the cluster and under CloudWatch metrics, observe Master Link Health Status. Once the the value reaches 1, the data is in sync.
- You can check that for the ElastiCache replica has an `online` state by running the `INFO` command on your Redis on EC2 instance. Doing this also provides information about replication lag.
- Verify low client output buffer by using the `CLIENT LIST` Redis command on your Redis on EC2 instance.
After the migration is complete, the ElastiCache cluster shows the status of **in-sync**. This status means that all data is now replicated. The data is in sync with any new writes coming to the master node of your Redis instance.

### Completing the Data Migration

When you are ready to cut over to the ElastiCache cluster, use the `complete-migration` CLI command with the following parameters:

- `--replication-group-id` – The identifier for the replication group.
- `--force` – A value that forces the migration to stop without ensuring that data is in sync.

The following is an example.

```
aws elasticache complete-migration --replication-group-id test-cluster
```

As you run this command, the ElastiCache primary node stops replicating from your Redis instance and promotes it to primary. This promotion typically completes within minutes. To confirm the promotion to primary, check for the event `Complete Migration successful for test-cluster`. At this point, you can direct your application to ElastiCache writes and reads. ElastiCache cluster status should change from **migrating** to **available**.

If the promotion to master fails, the ElastiCache primary node continues to replicate from your Redis on EC2 instance. The ElastiCache cluster continues to be in **migrating** status, and a replication group event message about the failure is sent. To troubleshoot this failure, look at the following:

- Check the replication group event. Use specific information from the event to fix the failure.
- You might get an event message about data not in sync. If so, make sure that the ElastiCache primary can replicate from your Redis on EC2 instance and both are in sync. If you still want to stop the migration, you can run the preceding command with the `--force` option.
- You might get an event message is about a node under replacement. If so, one of the ElastiCache nodes was undergoing a replacement. You can retry the complete migration step after the replacement is complete.

### Performing Online Data Migration Using the Console

You can use the AWS Management Console to migrate your data from the EC2 instance to your Redis cluster.

**To perform online data migration using the console**

2. Either create a new Redis cluster or choose an existing cluster. Make sure that the cluster meets the following requirements:
   - Your Redis engine version should be at least 5.0.5 (or higher).
   - Your Redis cluster should be in cluster-mode disabled configuration.
• You Redis on EC2 instance should not have Redis AUTH enabled.
• Redis config protected-mode should be set to no.
• If you have bind configuration in your Redis config, then it should be updated to allow requests from ElastiCache nodes.
• The number of databases should be the same between the ElastiCache node and your Redis on EC2 instance. This value is set using databases in the Redis config.
• Redis commands that perform data modification should not be renamed to allow replication of the data to succeed.
• To replicate the data from your Redis cluster to ElastiCache, make sure that there is sufficient CPU and memory to handle this additional load. This load comes from the RDB file created by your Redis cluster and transferred over the network to ElastiCache node.
• The cluster is in available status.

3. With your cluster selected, choose Migrate Data from Endpoint for Actions.
4. In the Migrate Data from Endpoint dialog box, enter either the IP address or the name of the EC2 instance, and the port where your Redis on EC2 instance is available.

   **Important**
   The IP address must be exact. If you enter the address incorrectly, the migration fails.

5. Choose Start Migration.

   As the cluster begins migration, it changes to Modifying and then Migrating status.

6. Monitor the migration progress by choosing Events on the navigation pane.
At any point during the migration process, you can stop migration. To do so, choose your cluster and choose **Stop Data Migration** for Actions. The cluster then goes to **Available** status.

If the migration succeeds, the cluster goes to **Available** status and the event log shows the following:

Migration operation succeeded for replication group `ElastiCacheClusterName`.

If the migration fails, the cluster goes to **Available** status and the event log shows the following:

Migration operation failed for replication group `ElastiCacheClusterName`. 
Managing Your ElastiCache for Redis 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. 44)
- Choosing Regions and Availability Zones (p. 55)
- Managing Maintenance (p. 59)
- Managing Nodes (p. 60)
- Managing Your ElastiCache Clusters (p. 72)
- Accessing Your Cluster or Replication Group (p. 119)
- Working with Shards (p. 127)
- Replication Across AWS Regions Using Global Datastore (p. 130)
- High Availability Using Replication Groups (p. 142)
- Finding Connection Endpoints (p. 204)
- Backup and Restore for ElastiCache for Redis (p. 213)
- Scaling ElastiCache for Redis Clusters (p. 250)
- Configuring Engine Parameters Using Parameter Groups (p. 298)

Engine Versions and Upgrading

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

Topics
- Supported ElastiCache for Redis Versions (p. 45)
- Upgrading Engine Versions (p. 53)
Supported ElastiCache for Redis Versions

You can use Amazon ElastiCache for Redis to build HIPAA-compliant applications. To help do this, you can enable at-rest encryption, in-transit encryption, and Redis AUTH when you create a Redis cluster using ElastiCache for Redis versions 3.2.6, 4.0.10, or later. You can store healthcare-related information, including protected health information (PHI), under an executed Business Associate Agreement (BAA) with AWS. AWS Services in Scope have been fully assessed by a third-party auditor and result in a certification, attestation of compliance, or Authority to Operate (ATO). For more information, see the following topics:

- AWS Cloud Compliance
- HIPAA Compliance
- AWS Services in Scope by Compliance Program
- ElastiCache for Redis Compliance (p. 450)
- Data Security in Amazon ElastiCache (p. 340)
- Authenticating Users with the Redis AUTH Command (p. 354)

Supported ElastiCache for Redis versions

- ElastiCache for Redis Version 5.0.6 (Enhanced) (p. 46)
- ElastiCache for Redis Version 5.0.5 (Enhanced) (p. 46)
- ElastiCache for Redis Version 5.0.4 (Enhanced) (p. 46)
- ElastiCache for Redis Version 5.0.3 (Enhanced) (p. 47)
- ElastiCache for Redis Version 5.0.0 (Enhanced) (p. 48)
- ElastiCache for Redis Version 4.0.10 (Enhanced) (p. 48)
- ElastiCache for Redis Version 3.2.10 (Enhanced) (p. 49)
- ElastiCache for Redis Version 3.2.6 (Enhanced) (p. 49)
- ElastiCache for Redis Version 3.2.4 (Enhanced) (p. 50)
- ElastiCache for Redis Version 2.8.24 (Enhanced) (p. 51)
- ElastiCache for Redis Version 2.8.23 (Enhanced) (p. 51)
- ElastiCache for Redis Version 2.8.22 (Enhanced) (p. 51)
- ElastiCache for Redis Version 2.8.21 (p. 52)
- ElastiCache for Redis Version 2.8.19 (p. 52)
- ElastiCache for Redis Version 2.8.6 (p. 52)
- ElastiCache for Redis Version 2.6.13 (p. 52)

**Note**
Because the newer Redis versions provide a better and more stable user experience, Redis versions 2.6.13, 2.8.6, and 2.8.19 are deprecated when using the ElastiCache console. We recommend against using these Redis versions. If you need to use one of them, work with the AWS CLI or ElastiCache API.

For more information, see the following topics:

<table>
<thead>
<tr>
<th></th>
<th>AWS CLI</th>
<th>ElastiCache API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Cluster</td>
<td>Creating a Cluster (AWS CLI) (p. 87)</td>
<td>Creating a Cluster (ElastiCache API) (p. 88)</td>
</tr>
</tbody>
</table>
Amazon ElastiCache for Redis introduces the next version of the Redis engine supported by Amazon ElastiCache, which includes bug fixes.

For more information, see Redis 5.0.6 Release Notes at Redis on GitHub.

Amazon ElastiCache for Redis introduces the next version of the Redis engine supported by Amazon ElastiCache. It includes online configuration changes for ElastiCache for Redis of auto-failover clusters during all planned operations. You can now scale your cluster, upgrade the Redis engine version and apply patches and maintenance updates while the cluster stays online and continues serving incoming requests. It also includes bug fixes.

For more information, see Redis 5.0.5 Release Notes at Redis on GitHub.

Amazon ElastiCache for Redis introduces the next version of the Redis engine supported by Amazon ElastiCache. It includes the following enhancements:

- Engine stability guarantee in special conditions.
- Improved Hyperloglog error handling.
- Enhanced handshake commands for reliable replication.
- Consistent message delivery tracking via XCLAIM command.
- Improved LFU field management in objects.
- Enhanced transaction management when using ZPOP.
Amazon ElastiCache for Redis introduces the next version of the Redis engine supported by Amazon ElastiCache. It includes the following enhancements:

- Bug fixes to improve sorted set edge cases, accurate memory usage and more. For more information, see Redis 5.0.3 release notes.
- Ability to rename commands: ElastiCache for Redis 5.0.3 includes a new parameter called `rename-commands` that allows you to rename potentially dangerous or expensive Redis commands that might cause accidental data loss, such as `FLUSHALL` or `FLUSHDB`. This is similar to the rename-command configuration in open source Redis. However, ElastiCache has improved the experience by providing a fully managed workflow. The command name changes are applied immediately, and automatically propagated across all nodes in the cluster that contain the command list. There is no intervention required on your part, such as rebooting nodes.

The following examples demonstrate how to modify existing parameter groups. They include the `rename-commands` parameter, which is a space-separated list of commands you want to rename:

```bash
code
aws elasticache modify-cache-parameter-group --cache-parameter-group-name custom_param_group --parameter-name-values "ParameterName=rename-commands, ParameterValue='flushall restrictedflushall'" --region region
```

In this example, the `rename-commands` parameter is used to rename the `flushall` command to `restrictedflushall`.

To rename multiple commands, use the following:

```bash
code
aws elasticache modify-cache-parameter-group --cache-parameter-group-name custom_param_group --parameter-name-values "ParameterName=rename-commands, ParameterValue='flushall restrictedflushall flushdb restrictedflushdb'" --region region
```

To revert any change, re-run the command and exclude any renamed values from the `ParameterValue` list that you want to retain, as shown following:

```bash
code
aws elasticache modify-cache-parameter-group --cache-parameter-group-name custom_param_group --parameter-name-values "ParameterName=rename-commands, ParameterValue='flushall restrictedflushall'" --region region
```

In this case, the `flushall` command is renamed to `restrictedflushall` and any other renamed commands revert to their original command names.

**Note**

When renaming commands, you are restricted to the following limitations:

- All renamed commands should be alphanumeric.
- The maximum length of new command names is 20 alphanumeric characters.
- When renaming commands, ensure that you update the parameter group associated with your cluster.
- To prevent a command's use entirely, use the keyword `blocked`, as shown following:

```bash
code
aws elasticache modify-cache-parameter-group --cache-parameter-group-name custom_param_group
```

API Version 2015-02-02
ElastiCache for Redis Version 5.0.0 (Enhanced)

Amazon ElastiCache for Redis introduces the next major version of the Redis engine supported by Amazon ElastiCache. ElastiCache for Redis 5.0.0 brings support for the following improvements:

- **Redis Streams:** This models a log data structure that allows producers to append new items in real time. It also allows consumers to consume messages either in a blocking or nonblocking fashion. Streams also allow consumer groups, which represent a group of clients to cooperatively consume different portions of the same stream of messages, similar to Apache Kafka. For more information, see Introduction to Redis Streams.
- **Support for a family of stream commands, such as `XADD`, `XRANGE` and `XREAD`**. For more information, see Redis Streams Commands.
- A number of new and renamed parameters. For more information, see Redis 5.0.0 Parameter Changes (p. 317).
- A new Redis metric, `StreamBasedCmds`.
- Slightly faster snapshot time for Redis nodes.

**Important**

Amazon ElastiCache for Redis has back-ported two critical bug fixes from Redis open source version 5.0.1. They are listed following:

- `RESTORE` mismatch reply when certain keys have already expired.
- The `XCLAIM` command can potentially return a wrong entry or desynchronize the protocol.

Both of these bug fixes are included in ElastiCache for Redis support for Redis engine version 5.0.0 and are consumed in future version updates.

ElastiCache for Redis Version 4.0.10 (Enhanced)

Amazon ElastiCache for Redis introduces the next major version of the Redis engine supported by Amazon ElastiCache. ElastiCache for Redis 4.0.10 brings support for the following improvements:

- Both online cluster resizing and encryption in a single ElastiCache for Redis version. For more information, see the following:
  - Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)
  - Online Resharding and Shard Rebalancing for Redis (cluster mode enabled) (p. 281)
  - Data Security in Amazon ElastiCache (p. 340)
- A number of new parameters. For more information, see Redis 4.0.10 Parameter Changes (p. 320).
- Support for family of memory commands, such as `MEMORY`. For more information, see Redis Commands (search on MEMO).
- Support for memory defragmentation while online thus allowing more efficient memory utilization and more memory available for your data.
- Support for asynchronous flushes and deletes. ElastiCache for Redis supports commands like `UNLINK`, `FLUSHDB` and `FLUSHALL` to run in a different thread from the main thread. Doing this helps improve performance and response times for your applications by freeing memory asynchronously.
A new Redis metric, `ActiveDefragHits`. For more information, see Metrics for Redis.

Redis (cluster mode disabled) users running Redis version 3.2.10 can use the console to upgrade their clusters via online upgrade.

**Comparing ElastiCache for Redis Cluster Resizing and Encryption Support**

<table>
<thead>
<tr>
<th>Feature</th>
<th>3.2.6</th>
<th>3.2.10</th>
<th>4.0.10 and later</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online cluster resizing *</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>In-transit encryption **</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>At rest encryption **</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* Adding, removing, and rebalancing shards.

** Required for FedRAMP, HIPAA, and PCI DSS compliant applications. For more information, see ElastiCache for Redis Compliance (p. 450).

**ElastiCache for Redis Version 3.2.10 (Enhanced)**

Amazon ElastiCache for Redis introduces the next major version of the Redis engine supported by Amazon ElastiCache. ElastiCache for Redis 3.2.10 introduces online cluster resizing to add or remove shards from the cluster while it continues to serve incoming I/O requests. ElastiCache for Redis 3.2.10 users have all the functionality of earlier Redis versions except the ability to encrypt their data. This ability is currently available only in version 3.2.6.

**Comparing ElastiCache for Redis versions 3.2.6 and 3.2.10**

<table>
<thead>
<tr>
<th>Feature</th>
<th>3.2.6</th>
<th>3.2.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online cluster resizing *</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>In-transit encryption **</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>At rest encryption **</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

* Adding, removing, and rebalancing shards.

** Required for FedRAMP, HIPAA, and PCI DSS compliant applications. For more information, see ElastiCache for Redis Compliance (p. 450).

For more information, see the following:

- Online Resharding and Shard Rebalancing for Redis (cluster mode enabled) (p. 281)
- Best Practices: Online Cluster Resizing (p. 472)

**ElastiCache for Redis Version 3.2.6 (Enhanced)**

Amazon ElastiCache for Redis introduces the next major version of the Redis engine supported by Amazon ElastiCache. ElastiCache for Redis 3.2.6 users have all the functionality of earlier Redis versions plus the option to encrypt their data. For more information, see the following:

- ElastiCache for Redis In-Transit Encryption (TLS) (p. 341)
ElastiCache for Redis Version 3.2.4 (Enhanced)

Amazon ElastiCache for Redis version 3.2.4 introduces the next major version of the Redis engine supported by Amazon ElastiCache. ElastiCache for Redis 3.2.4 users have all the functionality of earlier Redis versions available to them plus the option to run in *cluster mode* or *non-cluster mode*. The following table summarizes.

### Comparing Redis 3.2.4 Non-Cluster Mode and Cluster Mode

<table>
<thead>
<tr>
<th>Feature</th>
<th>Non-Cluster Mode</th>
<th>Cluster Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data partitioning</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Geospatial indexing</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Change node type</td>
<td>Yes</td>
<td>Yes *</td>
</tr>
<tr>
<td>Replica scaling</td>
<td>Yes</td>
<td>Yes *</td>
</tr>
<tr>
<td>Scale out</td>
<td>No</td>
<td>Yes *</td>
</tr>
<tr>
<td>Database support</td>
<td>Multiple</td>
<td>Single</td>
</tr>
<tr>
<td>Parameter group</td>
<td><code>default.redis3.2</code></td>
<td><code>default.redis3.2.cluster.on</code></td>
</tr>
</tbody>
</table>

* See [Restoring From a Backup with Optional Cluster Resizing](p. 239)

** Or one derived from it.

### Notes:

- **Partitioning** – the ability to split your data across 2 to 90 node groups (shards) with replication support for each node group.
- **Geospatial indexing** – Redis 3.2.4 introduces support for geospatial indexing via six GEO commands. For more information, see the Redis GEO* command documentation [Redis Commands: GEO](on the Redis Commands page (filtered for GEO).

For information about additional Redis 3 features, see [Redis 3.2 release notes](http://example.com) and [Redis 3.0 release notes](http://example.com).

Currently ElastiCache managed Redis (cluster mode enabled) does not support the following Redis 3.2 features:

- Replica migration
- Cluster rebalancing
- Lua debugger

ElastiCache disables the following Redis 3.2 management commands:

- `cluster meet`
- `cluster replicate`
- `cluster flushslots`
- cluster addslots
- cluster delslots
- cluster setslot
- cluster saveconfig
- cluster forget
- cluster failover
- cluster bumpepoch
- cluster set-config-epoch
- cluster reset

For information about Redis 3.2.4 parameters, see Redis 3.2.4 Parameter Changes (p. 323).

**ElastiCache for Redis Version 2.8.24 (Enhanced)**

Redis improvements added since version 2.8.23 include bug fixes and logging of bad memory access addresses. For more information, see Redis 2.8 release notes.

**ElastiCache for Redis Version 2.8.23 (Enhanced)**

Redis improvements added since version 2.8.22 include bug fixes. For more information, see Redis 2.8 release notes. This release also includes support for the new parameter `close-on-slave-write` which, if enabled, disconnects clients who attempt to write to a read-only replica.

For more information on Redis 2.8.23 parameters, see Redis 2.8.23 (Enhanced) Added Parameters (p. 326) in the ElastiCache User Guide.

**ElastiCache for Redis Version 2.8.22 (Enhanced)**

Redis improvements added since version 2.8.21 include the following:

- Support for forkless backups and synchronizations, which allows you to allocate less memory for backup overhead and more for your application. For more information, see How Synchronization and Backup are Implemented (p. 161). The forkless process can impact both latency and throughput. When there is high write throughput, when a replica re-syncs, it can be unreachable for the entire time it is syncing.
- If there is a failover, replication groups now recover faster because replicas perform partial syncs with the primary rather than full syncs whenever possible. Additionally, both the primary and replicas no longer use the disk during syncs, providing further speed gains.
- Support for two new CloudWatch metrics.
  - `ReplicationBytes` – The number of bytes a replication group's primary cluster is sending to the read replicas.
  - `SaveInProgress` – A binary value that indicates whether or not there is a background save process running.

  For more information, see Monitoring Use with Metrics (p. 413).
- A number of critical bug fixes in replication PSYNC behavior. For more information, see Redis 2.8 release notes.
- To maintain enhanced replication performance in Multi-AZ replication groups and for increased cluster stability, non-ElastiCache replicas are no longer supported.
- To improve data consistency between the primary cluster and replicas in a replication group, the replicas no longer evict keys independent of the primary cluster.
Redis configuration variables `appendonly` and `appendfsync` are not supported on Redis version 2.8.22 and later.

In low-memory situations, clients with a large output buffer might be disconnected from a replica cluster. If disconnected, the client needs to reconnect. Such situations are most likely to occur for PUBSUB clients.

**ElastiCache for Redis Version 2.8.21**

Redis improvements added since version 2.8.19 include a number of bug fixes. For more information, see Redis 2.8 release notes.

**ElastiCache for Redis Version 2.8.19**

Redis improvements added since version 2.8.6 include the following:

- Support for HyperLogLog. For more information, see Redis new data structure: HyperLogLog.
- The sorted set data type now has support for lexicographic range queries with the new commands `ZRANGEBYLEX`, `ZLEXCOUNT`, and `ZREMRANGEBYLEX`.
- To prevent a primary node from sending stale data to replica nodes, the master SYNC fails if a background save (`bgsave`) child process is aborted.
- Support for the `HyperLogLogBasedCommands` CloudWatch metric. For more information, see Metrics for Redis (p. 414).

**ElastiCache for Redis Version 2.8.6**

Redis improvements added since version 2.6.13 include the following:

- Improved resiliency and fault tolerance for read replicas.
- Support for partial resynchronization.
- Support for user-defined minimum number of read replicas that must be available at all times.
- Full support for pub/sub—notifying clients of events on the server.
- Automatic detection of a primary node failure and failover of your primary node to a secondary node.

**ElastiCache for Redis Version 2.6.13**

Redis version 2.6.13 was the initial version of Redis supported by Amazon ElastiCache for Redis. Multi-AZ with automatic failover is not supported on Redis 2.6.13.
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.

You initiate engine version upgrades to your cluster or replication group by modifying it and specifying a new engine version. For more information, see the following:

- Modifying an ElastiCache Cluster (p. 100)
- Modifying a Replication Group (p. 188)

**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).
- ElastiCache for Redis doesn't support switching between Redis (cluster mode disabled) and Redis (cluster mode enabled).
- The Amazon ElastiCache for Redis engine upgrade process is designed to make a best effort to retain your existing data and requires successful Redis replication.
- You can’t upgrade directly from Redis (cluster mode disabled) to Redis (cluster mode enabled) when you upgrade your engine. The following procedure shows you how to upgrade from Redis (cluster mode disabled) to Redis (cluster mode enabled).

**To upgrade from a Redis (cluster mode disabled) to Redis (cluster mode enabled) engine version**

1. Make a backup of your Redis (cluster mode disabled) cluster or replication group. For more information, see Making Manual Backups (p. 218).
2. Use the backup to create and seed a Redis (cluster mode enabled) cluster with one shard (node group). Specify the new engine version and enable cluster mode when creating the cluster or replication group. For more information, see Seeding a New Cluster with an Externally Created Backup (p. 242).
3. Delete the old Redis (cluster mode disabled) cluster or replication group. For more information, see Deleting a Cluster (p. 117) or Deleting a Replication Group (p. 190).
4. Scale the new Redis (cluster mode enabled) cluster or replication group to the number of shards (node groups) that you need. For more information, see Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)

- For single Redis clusters and clusters with Multi-AZ disabled, we recommend that sufficient memory be made available to Redis as described in Ensuring That You Have Enough Memory
to Create a Redis Snapshot (p. 464). In these cases, the primary is unavailable to service requests during the upgrade process.

- For Redis clusters with Multi-AZ enabled, we also recommend that you schedule engine upgrades during periods of low incoming write traffic. When upgrading to Redis 5.0.5 or above, the primary cluster continues to be available to service requests during the upgrade process. When upgrading to Redis 5.0.4 or below, you may notice a brief interruption of a few seconds associated with the DNS update.

Clusters and replication groups with multiple shards are process and patched as follows:

- All shards are processed in parallel. Only one upgrade operation is performed on a shard at any time.
- In each shard, all replicas are processed before the primary is processed. If there are fewer replicas in a shard, the primary in that shard might be processed before the replicas in other shards are finished processing.
- Across all the shards, primary nodes are processed in series. Only one primary node is upgraded at a time.
- If encryptions is enabled on your current cluster or replication group, you cannot upgrade to an engine version that does not support encryption, such as from 3.2.6 to 3.2.10.

### How to Upgrade Engine Versions

You initiate version upgrades to your cluster or replication group by modifying it using the ElastiCache console, the AWS CLI, or the ElastiCache API and specifying a newer engine version. For more information, see the following topics.

<table>
<thead>
<tr>
<th>How to Modify Clusters and Replication Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using the AWS Management Console (p. 100)</td>
</tr>
<tr>
<td>Using the AWS CLI (p. 101)</td>
</tr>
<tr>
<td>Using the ElastiCache API (p. 102)</td>
</tr>
</tbody>
</table>

### Resolving Blocked Redis Engine Upgrades

As shown in the following table, your Redis engine upgrade operation is blocked if you have a pending scale up operation.

<table>
<thead>
<tr>
<th>Pending Operations</th>
<th>Blocked Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale up</td>
<td>Immediate engine upgrade</td>
</tr>
<tr>
<td>Engine upgrade</td>
<td>Immediate scale up</td>
</tr>
<tr>
<td>Scale up and engine upgrade</td>
<td>Immediate scale up</td>
</tr>
<tr>
<td></td>
<td>Immediate engine upgrade</td>
</tr>
</tbody>
</table>

**To resolve a blocked Redis engine upgrade**

- Do one of the following:
• Schedule your Redis engine upgrade operation for the next maintenance window by clearing the Apply immediately check box.

  With the CLI, use --no-apply-immediately. With the API, use ApplyImmediately=false.

• Wait until your next maintenance window (or after) to perform your Redis engine upgrade operation.

• Add the Redis scale up operation to this cluster modification with the Apply Immediately check box chosen.

  With the CLI, use --apply-immediately. With the API, use ApplyImmediately=true.

  This approach effectively cancels the engine upgrade during the next maintenance window by performing it immediately.

### 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. 56).

**Regions and Availability Zones**

**Topics**

• Supported Regions & Endpoints (p. 56)

• Locating Your Nodes (p. 59)
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. 55) 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 (N. Virginia) Region</td>
<td>elasticache.us-east-1.amazonaws.com</td>
<td>HTTPS</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 (Oregon) Region</td>
<td>elasticache.us-west-2.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>Canada (Central) Region</td>
<td>elasticache.ca-central-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>Asia Pacific (Mumbai) Region</td>
<td>elasticache.ap-south-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>Asia Pacific (Tokyo) Region</td>
<td>elasticache.ap-northeast-1.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>Asia Pacific (Seoul) Region</td>
<td>elasticache.ap-northeast-2.amazonaws.com</td>
<td>HTTPS</td>
</tr>
<tr>
<td>Asia Pacific (Osaka-Local) Region *</td>
<td>elasticache.ap-northeast-3.amazonaws.com</td>
<td>HTTPS</td>
</tr>
</tbody>
</table>
## Supported Regions & Endpoints

<table>
<thead>
<tr>
<th>Region Name/Region</th>
<th>Endpoint</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<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. 66).

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. 74)
- Modifying an ElastiCache Cluster (p. 100)
- Adding Nodes to a Cluster (p. 104)

## Managing Maintenance

Every cluster and replication group 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 or replication group, 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>
Region Code | Region Name | Region Maintenance Window
--- | --- | ---
us-gov-west-1 | AWS GovCloud (US) region | 06:00–14:00 UTC
us-west-1 | US West (N. California) Region | 06:00–14:00 UTC
us-west-2 | US West (Oregon) Region | 06:00–14:00 UTC

**Changing your Cluster's or Replication Group's Maintenance Window**

The maintenance window should fall at the time of lowest usage and thus might need modification from time to time. You can modify your cluster or replication group to specify a time range of up to 24 hours in duration during which any maintenance activities you have requested should occur. Any deferred or pending cluster modifications you requested occur during this time.

**More information**

For information on your maintenance window and node replacement, see the following:

- ElastiCache Maintenance—FAQ on maintenance and node replacement
- Replacing Nodes (p. 69)—Managing node replacement
- Modifying an ElastiCache Cluster (p. 100)—Changing a cluster's maintenance window
- Modifying a Replication Group (p. 188)—Changing a replication group's maintenance window

**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 or replication group 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, Redis (cluster mode enabled) deployments have a number of smaller nodes. In contrast, Redis (cluster mode disabled) deployments have fewer, larger nodes in a cluster. For a more detailed discussion of which node size to use, see Choosing Your Node Size (p. 76).

**Topics**

- Redis Nodes and Shards (p. 61)
- Connecting to Nodes (p. 62)
- ElastiCache Reserved Nodes (p. 64)
- Supported Node Types (p. 65)
- Replacing Nodes (p. 69)

Some important operations involving nodes are the following:

- Adding Nodes to a Cluster (p. 104)
- Removing Nodes from a Cluster (p. 110)
- Scaling ElastiCache for Redis Clusters (p. 250)
- Finding Connection Endpoints (p. 204)
Redis Nodes and Shards

A shard (in the API and CLI, a node group) is a hierarchical arrangement of nodes, each wrapped in a cluster. Shards support replication. Within a shard, one node functions as the read/write primary node. All the other nodes in a shard function as read-only replicas of the primary node. Redis version 3.2 and later support multiple shards within a cluster (in the API and CLI, a replication group). This support enables partitioning your data in a Redis (cluster mode enabled) cluster.

The following diagram illustrates the differences between a Redis (cluster mode disabled) cluster and a Redis (cluster mode enabled) cluster.

![Diagram of Redis Clusters](image)

Both Redis (cluster mode disabled) and Redis (cluster mode enabled) support replication via shards. The API operation, DescribeReplicationGroups (CLI: describe-replication-groups) lists the node groups with the member nodes, the node’s role within the node group as well as other information.

When you create a Redis cluster, you specify whether you want to create a cluster with clustering enabled. Redis (cluster mode disabled) clusters never have more than one shard, which can be scaled horizontally by adding (up to a total of five) or deleting read replica nodes. For more information, see High Availability Using Replication Groups (p. 142), Adding a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 200) or Deleting a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 202). Redis (cluster mode disabled) clusters can also scale vertically by changing node types. For more information, see Scaling Redis (Cluster Mode Disabled) Clusters with Replica Nodes (p. 264).

When you create a Redis (cluster mode enabled) cluster, you specify from 1 to 90 shards.

**Note**
The node or shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see AWS Service Limits and choose the limit type **Nodes per cluster per instance type**.

After a Redis (cluster mode enabled) cluster is created, it can be altered (scaled in or out). For more information, see Scaling ElastiCache for Redis Clusters (p. 250) and Replacing Nodes (p. 69).

When you create a new cluster, you can seed it with data from the old cluster so it doesn’t start out empty. This approach works only if the cluster group has the same number of shards as the old cluster. Doing this can be helpful if you need change your node type or engine version. For more information, see Making Manual Backups (p. 218) and Restoring From a Backup with Optional Cluster Resizing (p. 239).
Connecting to Nodes

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

- Finding a Redis (Cluster Mode Disabled) Cluster's Endpoints (Console) (p. 206)
- Finding Endpoints for a Redis (Cluster Mode Enabled) Cluster (Console) (p. 208)
- Finding Endpoints (AWS CLI) (p. 210)
- Finding Endpoints (ElastiCache API) (p. 213)

In the following example, you use the `redis-cli` utility to connect to a cluster that is running Redis.

**Note**
For more information about Redis and available Redis commands, see the http://redis.io/commands webpage.

**To connect to a Redis cluster using the `redis-cli`**

1. Connect to your Amazon EC2 instance 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. To build `redis-cli`, download and install the GNU Compiler Collection (`gcc`). At the command prompt of your EC2 instance, enter the following command and enter y at the confirmation prompt.

   ```
   sudo yum install gcc
   ```

   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: 27 M
   Installed size: 53 M
   Is this ok [y/N]: y
   Downloading Packages:
   (1/11): binutils-2.22.52.0.1-10.36.amzn1.x86_64.rpm | 5.2 MB 00:00
   (2/11): cpp46-4.6.3-2.67.amzn1.x86_64.rpm | 4.8 MB 00:00
   (3/11): gcc-4.6.3-3.10.amzn1.noarch.rpm | 2.8 kB 00:00
   ...(output omitted)...
   Complete!
   ```

3. Download and compile the `redis-cli` utility. This utility is included in the Redis software distribution. At the command prompt of your EC2 instance, type the following commands:

   **Note**
   For Ubuntu systems, before running `make`, run `make distclean`.

   ```
   wget http://download.redis.io/redis-stable.tar.gz
   ```
4. At the command prompt of your EC2 instance, type the following command, substituting the endpoint of your cluster for the one shown in this example. Repeat this step for each node in your cluster that you want to connect to.

```
src/redis-cli -c -h mycachecluster.eaogs8.0001.usw2.cache.amazonaws.com -p 6379
```

A Redis command prompt similar to the following appears.

```
redis mycachecluster.eaogs8.0001.usw2.cache.amazonaws.com 6379>
```

5. Test the connection by running Redis commands.

You are now connected to the cluster and can run Redis commands. The following are some example commands with their Redis responses.

```
set a "hello"          // Set key "a" with a string value and no expiration
OK
get a                  // Get value for key "a"
"hello"
get b                  // Get value for key "b" results in miss
(nil)
set b "Good-bye" EX 5  // Set key "b" with a string value and a 5 second expiration
get b                  // wait 5 seconds
"Good-bye"
get b                  // key has expired, nothing returned
(nil)
quit                   // Exit from redis-cli
```

For connecting to nodes or clusters which have Secure Sockets Layer (SSL) encryption (in-transit enabled), see ElastiCache for Redis In-Transit Encryption (TLS) (p. 341).
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. 439).
## 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](https://docs.aws.amazon.com/AmazonElastiCache/latest/redис-guide/burstable-instance.html) in the Amazon ElastiCache 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>Node Type</th>
<th>CPU Credits Earned Per Hour</th>
<th>Maximum Earned Credits That Can Be Accrued*</th>
<th>vCPUs</th>
<th>Baseline Performance Per vCPU</th>
<th>Memory (GiB)</th>
<th>Network Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>t3.micro</td>
<td>12</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>24</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>24</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.
- Redis append-only files (AOF) aren't supported for T1 or T2 instances.
- Redis Multi-AZ with automatic failover isn't supported on T1 instances.
- Redis configuration variables appendonly and appendfsync aren't supported on Redis version 2.8.22 and later.

**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 Redis Versions](#).

### 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 East (Ohio)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<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></td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Europe-west-1 (Stockholm)</td>
<td>Yes</td>
<td></td>
<td></td>
<td>Yes</td>
<td>No</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>AWS Region</td>
<td>T3</td>
<td>T2</td>
<td>M4</td>
<td>M5</td>
<td>R4</td>
<td>R5</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Europe-west-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-west-2 (London)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>EU-west-3 (Paris)</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Southeast-1 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></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></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Middle-east-1 (Bahrain)</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>AWS-gov-west-1 (GovCloud (US-West)</td>
<td>Yes</td>
<td>Yes</td>
<td></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
- Redis Specific Parameters (p. 315)
Replacing Nodes

Amazon ElastiCache for Redis 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 Redis 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 is to take place. If you've 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 Redis nodes for replacement. To expedite finding the information you need for your situation, choose from the following menu.

- **Do nothing** (p. 69) – Let Amazon ElastiCache replace the node as scheduled.
- **Change your maintenance window** (p. 70) – Change your maintenance window to a better time.
- **Redis (cluster mode enabled) Configurations**
  - Replace the only node in any Redis cluster (p. 70) – A procedure to replace a node in a Redis cluster using backup and restore.
  - Replace a replica node in any Redis cluster (p. 71) – A procedure to replace a read-replica in any Redis cluster by increasing and decreasing the replica count with no cluster downtime.
  - Replace any node in a Redis (cluster mode enabled) shard (p. 71) – A dynamic procedure with no cluster downtime to replace a node in a Redis (cluster mode enabled) cluster by scaling out and scaling in.
- **Redis (cluster mode disabled) Configurations**
  - Replace the only node in any Redis cluster (p. 70) – Procedure to replace any node in a Redis cluster using backup and restore.
  - Replace a replica node in any Redis cluster (p. 71) – A procedure to replace a read-replica in any Redis cluster by increasing and decreasing the replica count with no cluster downtime.
  - Replace a node in a Redis (cluster mode disabled) cluster (p. 71) – Procedure to replace a node in a Redis (cluster mode disabled) cluster using replication.
  - Replace a Redis (cluster mode disabled) read-replica (p. 71) – A procedure to manually replace a read-replica in a Redis (cluster mode disabled) replication group.
  - Replace a Redis (cluster mode disabled) primary node (p. 72) – A procedure to manually replace the primary node in a Redis (cluster mode disabled) replication group.

**Redis node replacement options**

- **Do nothing** – If you do nothing, ElastiCache replaces the node as scheduled.

If the node is a member of an auto failover enabled cluster, ElastiCache for Redis provides improved availability during patching, updates, and other maintenance-related node replacements.
For ElastiCache for Redis Cluster configurations that are set up to use ElastiCache for Redis Cluster clients, replacement now completes while the cluster serves incoming write requests.

For non-Cluster configurations with autofailover enabled, clusters on Redis 5.0.5 complete replacement while the cluster continues to stay online and serve incoming write requests. For auto failover enabled clusters on Redis 5.0.4 or below, you might notice a brief write interruption of up to a few seconds associated with DNS updates.

If the node is standalone, Amazon ElastiCache first launches a replacement node and then syncs from the existing node. The existing node isn’t available for service requests during this time. Once the sync is complete, the existing node is terminated and the new node takes its place. ElastiCache makes a best effort to retain your data during this operation.

- **Change your maintenance window** – For scheduled maintenance events, you receive an email or a notification event from ElastiCache. In these cases, if you change your maintenance window before the scheduled replacement time, your node now is replaced at the new time. For more information, see the following:
  - Modifying an ElastiCache Cluster (p. 100)
  - Modifying a Replication Group (p. 188)

  **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. 59).

- **Replace the only node in any Redis cluster** – If the cluster does not have any read replicas, you can use the following procedure to replace the node.

  **To replace the only node using backup and restore**
  1. Create a snapshot of the node's cluster. For instructions, see Making Manual Backups (p. 218).
  2. Create a new cluster seeding it from the snapshot. For instructions, see Restoring From a Backup with Optional Cluster Resizing (p. 239).
  3. Delete the cluster with the node scheduled for replacement. For instructions, see Deleting a Cluster (p. 117).
  4. In your application, replace the old node’s endpoint with the new node’s endpoint.
• **Replace a replica node in any Redis cluster** – To replace a replica cluster, increase your replica count. To do this, add a replica then decrease the replica count by removing the replica that you want to replace. This process is dynamic and doesn’t have any cluster downtime.

  **Note**
  If your shard or replication group already has five replicas, reverse steps 1 and 2.

**To replace a replica in any Redis cluster**

1. Increase the replica count by adding a replica to the shard or replication group. For more information, see [Increasing the Number of Replicas in a Shard](p. 192).
2. Delete the replica you want to replace. For more information, see [Decreasing the Number of Replicas in a Shard](p. 196).
3. Update the endpoints in your application.

• **Replace any node in a Redis (cluster mode enabled) shard** – To replace the node in a cluster with no downtime, use online resharding. First add a shard by scaling out, and then delete the shard with the node to be replaced by scaling in.

**To replace any node in a Redis (cluster mode enabled) cluster**

1. Scale out: Add an additional shard with the same configuration as the existing shard with the node to be replaced. For more information, see [Adding Shards with Online Resharding](p. 281).
2. Scale in: Delete the shard with the node to be replaced. For more information, see [Removing Shards with Online Resharding](p. 284).
3. Update the endpoints in your application.

• **Replace a node in a Redis (cluster mode disabled) cluster** – If the cluster is a Redis (cluster mode disabled) cluster without any read replicas, use the following procedure to replace the node.

**To replace the node using replication (cluster mode disabled only)**

1. Add replication to the cluster with the node scheduled for replacement as the primary. Do not enable Multi-AZ on this cluster. For instructions, see [To add replication to a Redis cluster with no shards](p. 105).
2. Add a read-replica to the cluster. For instructions, see [To add nodes to a cluster (console)](p. 105).
3. Promote the newly created read-replica to primary. For instructions, see [Promoting a Read Replica to Primary, for Redis (cluster mode disabled) Replication Groups](p. 203).
4. Delete the node scheduled for replacement. For instructions, see [Removing Nodes from a Cluster](p. 110).
5. In your application, replace the old node's endpoint with the new node's endpoint.

• **Replace a Redis (cluster mode disabled) read-replica** – If the node is a read-replica, replace the node.

If your cluster has only one replica node and Multi-AZ is enabled, you must disable Multi-AZ before you can delete the replica. For instructions, see [Modifying a Replication Group](p. 188).
To replace a Redis (cluster mode disabled) read replica

1. Delete the replica that is scheduled for replacement. For instructions, see the following:
   - Decreasing the Number of Replicas in a Shard (p. 196)
   - Removing Nodes from a Cluster (p. 110)

2. Add a new replica to replace the one that is scheduled for replacement. If you use the same name as the replica you just deleted, you can skip step 3. For instructions, see the following:
   - Increasing the Number of Replicas in a Shard (p. 192)
   - Adding a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 200)

3. In your application, replace the old replica's endpoint with the new replica's endpoint.

4. If you disabled Multi-AZ at the start, re-enable it now. For instructions, see Enabling Multi-AZ with Automatic Failover (p. 154).

Replace a Redis (cluster mode disabled) primary node – If the node is the primary node, first promote a read-replica to primary. Then delete the replica that used to be the primary node.

If your cluster has only one replica and Multi-AZ is enabled, you must disable Multi-AZ before you can delete the replica in step 2. For instructions, see Modifying a Replication Group (p. 188).

To replace a Redis (cluster mode disabled) primary node

1. Promote a read-replica to primary. For instructions, see Promoting a Read Replica to Primary, for Redis (cluster mode disabled) Replication Groups (p. 203).

2. Delete the node that is scheduled for replacement (the old primary). For instructions, see Removing Nodes from a Cluster (p. 110).

3. Add a new replica to replace the one scheduled for replacement. If you use the same name as the node you just deleted, you can skip changing endpoints in your application.

   For instructions, see Adding a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 200).

4. In your application, replace the old node's endpoint with the new node's endpoint.

5. If you disabled Multi-AZ at the start, re-enable it now. For instructions, see Enabling Multi-AZ with Automatic Failover (p. 154).

Managing Your ElastiCache Clusters

A cluster is a collection of one or more cache nodes, all of which run an instance of the Redis 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 Redis cluster. Redis clusters can contain a single node or up to six nodes inside a shard (API/CLI: node group), A single-node Redis (cluster mode disabled) cluster has no shard, and a multi-node Redis (cluster mode disabled) cluster has a single shard. Redis (cluster mode enabled) clusters can have up to 90 shards, with your data partitioned across the shards. When you have multiple nodes in a shard, one of the nodes is a read/write primary node. All other nodes in the shard are read-only replicas.
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. 28)
- Accessing ElastiCache Resources from Outside AWS (p. 123)

**Supported Redis Versions**

- ElastiCache for Redis Version 5.0.0 (Enhanced) (p. 48)
- ElastiCache for Redis Version 4.0.10 (Enhanced) (p. 48)
- ElastiCache for Redis Version 3.2.10 (Enhanced) (p. 49)
- ElastiCache for Redis Version 3.2.6 (Enhanced) (p. 49)
- ElastiCache for Redis Version 3.2.4 (Enhanced) (p. 50)
- ElastiCache for Redis Version 2.8.23 (Enhanced) (p. 51)
- ElastiCache for Redis Version 2.8.22 (Enhanced) (p. 51)
- ElastiCache for Redis Version 2.8.19 (p. 52)
- ElastiCache for Redis Version 2.8.6 (p. 52)
- ElastiCache for Redis Version 2.6.13 (p. 52)

**Other ElastiCache Cluster Operations**

Additional operations involving clusters:

- Finding Connection Endpoints (p. 204)
- Accessing ElastiCache Resources from Outside AWS (p. 123)
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 or replication group, 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. 74)
• Choosing Your Node Size (p. 76)
• Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79)
• Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 83)
• Creating a Cluster (AWS CLI) (p. 87)
• Creating a Cluster (ElastiCache API) (p. 88)

Determine Your Requirements

Topics

• Memory and Processor Requirements (p. 75)
• Redis Cluster Configuration (p. 75)
• Scaling Requirements (p. 75)
• Access Requirements (p. 75)
• Region and Availability Zone Requirements (p. 76)

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 Node Size (p. 76).

• 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. 374).
  
  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. 123).

• 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. 300).

  If you're running Redis you may want to consider at least setting reserved-memory or reserved-memory-percent. For more information, see Managing Reserved Memory (p. 466).
Creating a Cluster

• Do you need to create your own Security Group or VPC Security Group?
  
  For more information, see Security Groups: EC2-Classic (p. 382) and Security in Your VPC.

• How do you intend to implement fault tolerance?
  
  For more information, see Mitigating Failures (p. 454).

Topics

• Memory and Processor Requirements (p. 75)
• Redis Cluster Configuration (p. 75)
• Scaling Requirements (p. 75)
• Access Requirements (p. 75)
• Region and Availability Zone Requirements (p. 76)

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.

Redis Cluster Configuration

ElastiCache for Redis clusters are comprised of from 0 to 90 shards (also called node groups). The data in a Redis cluster is partitioned across the shards in the cluster. Your application connects with a Redis cluster using a network address called an Endpoint. The nodes in a Redis shard fulfill one of two roles: one read/write primary and all other nodes read-only secondaries (also called read replicas). In addition to the node endpoints, the Redis 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 ElastiCache for Redis.

For more information, see Managing Your ElastiCache Clusters (p. 72).

Scaling Requirements

All clusters can be scaled up by creating a new cluster with the new, larger node type. When scaling a Redis cluster you can seed it from a backup and avoid having the new cluster start out empty.

For more information, see Scaling ElastiCache for Redis Clusters (p. 250) 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. 28) 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. 55)
- Mitigating Failures (p. 454)

Choosing Your Node Size

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

Choosing Your Node Size

Answering the following questions can help you determine the minimum node type you need for your Redis implementation:

- How much total memory do you need for your data?

To get a general estimate, take the size of the items that you want to cache. Multiply this size by the number of items that you want to keep in the cache at the same time. To get a reasonable estimation of the item size, first serialize your cache items, then count the characters. Then divide this over the number of shards in your cluster.

- What version of Redis are you running?

Redis versions before 2.8.22 require you to reserve more memory for failover, snapshot, synchronizing, and promoting a replica to primary operations. This requirement occurs because you must have sufficient memory available for all writes that occur during the process.

Redis version 2.8.22 and later use a forkless save process that requires less available memory than the earlier process.

For more information, see the following:

- How Synchronization and Backup are Implemented (p. 161)
- Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464)

- How write-heavy is your application?
Write heavy applications can require significantly more available memory, memory not used by data, when taking snapshots or failing over. Whenever the BGSAVE process is performed, you must have sufficient memory that is unused by data to accommodate all the writes that transpire during the BGSAVE process. Examples are when taking a snapshot, when syncing a primary cluster with a replica in a cluster, and when enabling the append-only file (AOF) feature. Another is when promoting a replica to primary (if you have Multi-AZ with auto failover enabled). The worst case is when all of your data is rewritten during the process. In this case, you need a node instance size with twice as much memory as needed for data alone.

For more detailed information, see Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464).

- Will your implementation be a standalone Redis (cluster mode disabled) cluster or a Redis (cluster mode enabled) cluster with multiple shards?

**Redis (cluster mode disabled) cluster**

If you’re implementing a Redis (cluster mode disabled) cluster, your node type must be able to accommodate all your data plus the necessary overhead as described in the previous bullet.

For example, suppose that you estimate that the total size of all your items is 12 GB. In this case, you can use a cache.m3.xlarge node with 13.3 GB of memory or a cache.r3.large node with 13.5 GB of memory. However, you might need more memory for BGSAVE operations. If your application is write-heavy, double the memory requirements to at least 24 GB. Thus, use either a cache.m3.2xlarge with 27.9 GB of memory or a cache.r3.xlarge with 30.5 GB of memory.

**Redis (cluster mode enabled) with multiple shards**

If you're implementing a Redis (cluster mode enabled) cluster with multiple shards, then the node type must be able to accommodate bytes-for-data-and-overhead / number-of-shards bytes of data.

For example, suppose that you estimate that the total size of all your items to be 12 GB and you have two shards. In this case, you can use a cache.m3.large node with 6.05 GB of memory (12 GB / 2). However, you might need more memory for BGSAVE operations. If your application is write-heavy, double the memory requirements to at least 12 GB per shard. Thus, use either a cache.m3.xlarge with 13.3 GB of memory or a cache.r3.large with 13.5 GB of memory.

While your cluster is running, you can monitor the memory usage, processor utilization, cache hits, and cache misses metrics that are published to CloudWatch. You might notice that your cluster doesn’t have the hit rate that you want or that keys are being evicted too often. In these cases, you can choose a different node size with larger CPU and memory specifications.
When monitoring CPU usage, remember that Redis is single-threaded. Thus, multiply the reported CPU usage by the number of CPU cores to get that actual usage. For example, a four-core CPU reporting a 20-percent usage rate is actually the one core Redis is using running at 80 percent.
Creating a Redis (cluster mode disabled) Cluster (Console)

ElastiCache supports replication when you use the Redis engine. To monitor the latency between when data is written to a Redis read/write primary cluster and when it is propagated to a read-only secondary cluster, ElastiCache adds to the cluster a special key, ElastiCacheMasterReplicationTimestamp. This key is the current Universal Coordinated Time (UCT) time. Because a Redis cluster might be added to a replication group at a later time, this key is included in all Redis clusters, even if initially they are not members of a replication group. For more information on replication groups, see High Availability Using Replication Groups (p. 142).

To create a standalone Redis (cluster mode disabled) cluster

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 Redis from the navigation pane.
4. Choose Create.
5. For Cluster engine, choose Redis, and then clear the Cluster Mode enabled (Scale Out) check box.
6. Complete the Redis settings section.
   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. In the Description box, type in a description for this cluster.
   c. For Engine version compatibility, choose the ElastiCache for Redis 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.

   Important
   You can upgrade to newer engine versions. For more information on doing so, see Upgrading Engine Versions (p. 53). However, you can't downgrade to older engine versions except by deleting the existing cluster and creating it again.

Because the newer Redis versions provide a better and more stable user experience, Redis versions 2.6.13, 2.8.6, and 2.8.19 are deprecated when using the ElastiCache console. We recommend against using these Redis versions. If you need to use one of them, work with the AWS CLI or ElastiCache API.

For more information, see the following topics:

<table>
<thead>
<tr>
<th></th>
<th>AWS CLI</th>
<th>ElastiCache API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Cluster</td>
<td>Creating a Cluster (AWS CLI) (p. 87)</td>
<td>Creating a Cluster (ElastiCache API) (p. 88)</td>
</tr>
<tr>
<td></td>
<td>You can't use this action to create a replication group</td>
<td>You can't use this action to create a replication group with</td>
</tr>
<tr>
<td></td>
<td>with cluster mode enabled.</td>
<td>cluster mode enabled.</td>
</tr>
<tr>
<td>Modify Cluster</td>
<td>Using the AWS CLI (p. 101)</td>
<td>Using the ElastiCache API (p. 102)</td>
</tr>
</tbody>
</table>
You can't use this action to create a replication group with cluster mode enabled.

### Create Replication Group

<table>
<thead>
<tr>
<th>AWS CLI</th>
<th>ElastiCache API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch (AWS CLI) (p. 169)</td>
<td>Creating a Redis (cluster mode disabled) Replication Group from Scratch (ElastiCache API) (p. 172)</td>
</tr>
<tr>
<td>Creating a Redis (Cluster Mode Enabled) Replication Group from Scratch (AWS CLI) (p. 175)</td>
<td>Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (ElastiCache API) (p. 179)</td>
</tr>
</tbody>
</table>

### Modify Replication Group

<table>
<thead>
<tr>
<th>AWS CLI</th>
<th>ElastiCache API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using the AWS CLI (p. 188)</td>
<td>Using the ElastiCache API (p. 189)</td>
</tr>
</tbody>
</table>

**d.** To encrypt your data while it is in transit, for **Encryption**, choose **Yes**.

**e.** If you chose **Yes** for **Encryption**, you can require users to enter a password when executing Redis commands. To require a password when executing commands, do the following:

1. Choose **Yes** from the **AUTH** list.
2. Type in a password in the **AUTH token** box:

**f.** In **Port**, accept the default port, 6379. If you have a reason to use a different port, type the port number.

**g.** For **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 runtime parameters of your cluster. For more information on parameter groups, see Redis Specific Parameters (p. 315) and Creating a Parameter Group (p. 300).

**h.** 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 Node Size (p. 76).

**i.** For **Number of replicas**, choose the number of read replicas you want for this cluster.

If you choose **None**, the **description** and Multi-AZ with Auto-Failover fields disappear and the cluster your create look like the following.

**Redis (cluster mode disabled) cluster created with no replica nodes**

If you choose one or more replicas, the cluster you create looks something like the following.
7. Choose Advanced Redis settings and complete the section.

   a. If you chose to have one or more replicas, the Multi-AZ with Auto-Failover check box is available. We strongly suggest that you enable Multi-AZ with Auto-Failover. For more information, see Mitigating Failures when Running Redis (p. 454).
   
   b. For Subnet group, choose the subnet you want to apply to this cluster.

   For more information, see Subnets and Subnet Groups (p. 374).

   c. For Availability zone(s), you have two options:
      
      • No preference – ElastiCache chooses the Availability Zones for your cluster's nodes.
      • Specify availability zones – A list of your nodes appears allowing you to specify the Availability Zone for each node in your cluster by choosing the Availability Zone from the list to the right of each node name.

   For more information, see Choosing Regions and Availability Zones (p. 55).

   d. For Security groups, choose the security groups you want for this cluster.

   For more information, see Amazon VPCs and ElastiCache Security (p. 358).

   e. If you are going to seed your cluster with data from a .RDB file, in the Seed RDB file S3 location box, type the Amazon S3 location of the .RDB file.

   For more information, see Seeding a New Cluster with an Externally Created Backup (p. 242).

   f. If you want regularly scheduled automatic backups, choose Enable automatic backups, and then type the number of days you want an automatic backup retained before it is automatically deleted. If you don't want regularly scheduled automatic backups, clear the Enable automatic backups check box. In either case, you always have the option to create manual backups, which must be deleted manually.

   For more information on Redis backup and restore, see Backup and Restore for ElastiCache for Redis (p. 213).

   g. 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 to 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. 59).

   h. 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.

API Version 2015-02-02

81
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. 28) and Step 3: Connect to a Cluster's Node (p. 29).

**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. 117).
Creating a Redis (Cluster Mode Enabled) Cluster (Console)

If you are running Redis 3.2.4 or later, you can create a Redis (cluster mode enabled) cluster. Redis (cluster mode enabled) clusters support partitioning your data across 1 to 90 shards (API/CLI: node groups) but with some limitations. For a comparison of Redis (cluster mode disabled) and Redis (cluster mode enabled), see Supported ElastiCache for Redis Versions (p. 45).

You can create a Redis (cluster mode enabled) cluster (API/CLI: replication group) using the ElastiCache management console, the AWS CLI for ElastiCache, and the ElastiCache API.

To create a Redis (cluster mode enabled) 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 Redis from the navigation pane.
4. Choose Create.
5. For Cluster engine, choose Redis, and then choose Cluster Mode enabled (Scale Out). These selections create a Redis (cluster mode enabled) cluster that looks something like the following.

   Redis (cluster mode enabled) cluster created with replication and data partitioning

6. Complete the Redis (cluster mode enabled) settings section.
   a. In the Name box, 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. In the Description box, type a description of the cluster.
   c. If you want to enable in-transit encryption for this cluster, choose In-transit encryption.
      If you choose In-transit encryption, two additional options appear: Redis auth token and a box where you type in the token (password) value.
   d. If you want to enable at-rest encryption for this cluster, choose At-rest encryption.
   e. To require a password for operations to be performed on this cluster:
      i. Choose Redis auth token.
      ii. In the Redis auth token box, type the token (password) that must be used when performing operations on this cluster.
   f. For Engine version compatibility, choose the ElastiCache for Redis 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.
   g. To encrypt your data while it is in transit, for Encryption, choose Yes.
h. If you chose Yes for Encryption, you can require users to enter a password when executing Redis commands. To require a password when executing commands, do the following:

i. Choose Yes from the AUTH list.
ii. Type in a password in the AUTH token box.

i. In the Port box, accept the default port, 6379. If you have a reason to use a different port, type the port number.

j. For 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 Redis Specific Parameters (p. 315) and Creating a Parameter Group (p. 300).

k. For Node type, choose 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 Node Size (p. 76).

l. For Number of shards, choose the number of shards (partitions/node groups) you want for this Redis (cluster mode enabled) cluster.

In Redis (cluster mode enabled), depending upon the version of Redis running on your cluster, you may be able to change the number of shards in your cluster dynamically.

- **Redis 3.2.10**—If your cluster is running Redis 3.2.10 you can change the number of shards in your cluster dynamically. For more information, see Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279).

- **Other Redis versions**—If your cluster is running a version of Redis other than version 3.2.10, to change the number of shards in your cluster, you must create a new cluster with the new number of shards. For more information, see Restoring From a Backup with Optional Cluster Resizing (p. 239).

m. For Replicas per shard, choose the number of read replica nodes you want in each shard.

The following restrictions exist for Redis (cluster mode enabled).

- The number of replicas is the same for each shard when creating the cluster using the console.

- The number of read replicas per shard is fixed and cannot be changed. If you find you need more or fewer replicas per shard (API/CLI: node group), you must create a new cluster with the new number of replicas. For more information, see Seeding a New Cluster with an Externally Created Backup (p. 242).

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

For more information, see Subnets and Subnet Groups (p. 374).

7. Click Advanced Redis settings and complete the section.

a. For Slots and keyspaces, choose how you want your keys distributed over your shards (partitions). There are 16,384 keys to be distributed (numbered 0 through 16383).

- **Equal distribution** – ElastiCache distributes your keyspace as equally as possible over your shards.
Creating a Cluster

- **Custom distribution** – You specify the range of keys for each shard in the table below Availability zone(s).

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

- **No preference** – ElastiCache chooses the Availability Zone.

- **Specify availability zones** – You specify the Availability Zone for each cluster.

If you chose to specify the Availability Zones, for each cluster in each shard, choose the Availability Zone from the list.

For more information, see Choosing Regions and Availability Zones (p. 55).

Specifying Keyspaces and Availability Zones
c. For **Security groups**, choose the security groups you want for this cluster.

For more information, see Amazon VPCs and ElastiCache Security (p. 358).

d. If you are going to seed your cluster with data from a .RDB file, in the Seed RDB file S3 location box, enter the S3 location of the .RDB file.

For more information, see Seeding a New Cluster with an Externally Created Backup (p. 242).

For Redis (cluster mode enabled) you must have a separate .RDB file for each node group.

e. If you want regularly scheduled automatic backups, choose **Enable automatic backups** the type the number of days you want each automatic backup retained before it is automatically deleted. If you don't want regularly scheduled automatic backups, clear the **Enable automatic backups** check box. In either case, you always have the option to create manual backups.

For more information on Redis backup and restore, see Backup and Restore for ElastiCache for Redis (p. 213).

f. 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 to 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 from the lists, 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. 59).
9. For **Notifications**, choose an existing Amazon Simple Notification Service (Amazon SNS) topic, or choose Manual ARN input and type in the topic's 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 cluster** to launch your cluster, or **Cancel** to cancel the operation.

To create the equivalent using the ElastiCache API or AWS CLI instead of the ElastiCache console, see the following:

- API: CreateReplicationGroup
- CLI: create-replication-group

As soon as your cluster's status is **available**, you can grant EC2 access to it, connect to it, and begin using it. For more information, see Step 2: Authorize Access (p. 28) and Step 3: Connect to a Cluster's Node (p. 29).

**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. 117).
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. See Deleting a Cluster (p. 117).

**Topics**

- Creating a Cache Cluster for Redis (Cluster Mode Disabled) (AWS CLI) (p. 87)
- Creating a Redis (Cluster Mode Enabled) Cluster (AWS CLI) (p. 87)

Creating a Cache Cluster for Redis (Cluster Mode Disabled) (AWS CLI)

**Example – A Redis (cluster mode disabled) Cluster with no read replicas**

The following CLI code creates a Redis (cluster mode disabled) cache cluster with no replicas.

For Linux, macOS, or Unix:

```bash
aws elasticache create-cache-cluster \
  --cache-cluster-id my-cluster \
  --cache-node-type cache.r4.large \
  --engine redis \
  --engine-version 3.2.4 \
  --num-cache-nodes 1 \
  --cache-parameter-group default.redis3.2 \
  --snapshot-arns arn:aws:s3:myS3Bucket/snap.rdb
```

For Windows:

```bash
aws elasticache create-cache-cluster ^
  --cache-cluster-id my-cluster ^
  --cache-node-type cache.r4.large ^
  --engine redis ^
  --engine-version 3.2.4 ^
  --num-cache-nodes 1 ^
  --cache-parameter-group default.redis3.2 ^
  --snapshot-arns arn:aws:s3:myS3Bucket/snap.rdb
```

Creating a Redis (Cluster Mode Enabled) Cluster (AWS CLI)

Redis (cluster mode enabled) clusters (API/CLI: replication groups) cannot be created using the create-cache-cluster operation. To create a Redis (cluster mode enabled) cluster (API/CLI: replication group), see Creating a Redis (Cluster Mode Enabled) Replication Group from Scratch (AWS CLI) (p. 175).

For more information, see the AWS CLI for ElastiCache reference topic create-replication-group.)
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. 117).

**Topics**

- Creating a Redis (Cluster Mode Disabled) Cache Cluster (ElastiCache API) (p. 88)
- Creating a Cache Cluster in Redis (Cluster Mode Enabled) (ElastiCache API) (p. 88)

Creating a Redis (Cluster Mode Disabled) Cache Cluster (ElastiCache API)

The following code creates a Redis (cluster mode disabled) cache cluster (ElastiCache API).

```plaintext
https://elasticache.us-west-2.amazonaws.com/
?Action=CreateCacheCluster
&CacheClusterId=my-cluster
&CacheNodeType=cache.r4.large
&CacheParameterGroup=default.redis3.2
&Engine=redis
&EngineVersion=3.2.4
&NumCacheNodes=1
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&SnapshotArns.member.1=arn%3Aaws%3As3%3A%3A%3A%3AmyS3Bucket%2Fdump.rdb
&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>
```

Creating a Cache Cluster in Redis (Cluster Mode Enabled) (ElastiCache API)

Redis (cluster mode enabled) clusters (API/CLI: replication groups) cannot be created using the `CreateCacheCluster` operation. To create a Redis (cluster mode enabled) cluster (API/CLI: replication group), see Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (ElastiCache API) (p. 179).

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

API Version 2015-02-02

88
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 Redis (Cluster Mode Disabled) Cluster's Details (Console) (p. 90)
- Viewing Details for a Redis (Cluster Mode Enabled) Cluster (Console) (p. 91)
- Viewing a Cluster's Details (AWS CLI) (p. 92)
- Viewing a Cluster's Details (ElastiCache API) (p. 98)
Viewing a Redis (Cluster Mode Disabled) Cluster's Details (Console)

You can view the details of a Redis (cluster mode disabled) 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 Redis (cluster mode disabled) cluster using the ElastiCache console.

To view a Redis (cluster mode disabled) 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. In the ElastiCache console dashboard, choose Redis to display a list of all your clusters that are running any version of Redis.
3. To see details of a cluster, select the check box to the left of the cluster's name. Make sure you select a cluster running the Redis engine, not Clustered Redis. Doing this displays details about the cluster, including the cluster's primary endpoint.
4. To view node information:
   a. Choose the cluster's name.
   b. Choose the Nodes tab. Doing this displays details about each node, including the node's endpoint which you need to use to read from the cluster.
   c. To view metrics on one or more nodes, select the box to the left of the node ID, then select the time range for the metrics from the Time range list. If you select multiple nodes, you can see overlay graphs.
Viewing Details for a Redis (Cluster Mode Enabled) Cluster (Console)

You can view the details of a Redis (cluster mode enabled) 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 Redis (cluster mode enabled) cluster using the ElastiCache console.

To view a Redis (cluster mode enabled) 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 Redis to display a list of all your clusters that are running any version of Redis.
4. To see details of a Redis (cluster mode enabled) cluster, choose the box to the left of the cluster's name. Make sure you choose a cluster running the Clustered Redis engine, not just Redis.

   The screen expands below the cluster and display details about the cluster, including the cluster's configuration endpoint.

5. To see a listing of the cluster's shards and the number of nodes in each shard, choose the cluster's name.
6. To view specific information on a node:
   a. Choose the shard's ID.
   b. Choose the Nodes tab.

      This will display information about each node, including each node's endpoint that you need to use to read data from the cluster.
   c. To view metrics on one or more nodes, choose the box to the left of the node's id, and then choose the time range for the metrics from the Time range list. Selecting multiple nodes will generate overlay graphs.
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`.

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

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

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

Example

For Linux, macOS, or Unix:

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

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

The following JSON output is for three different Redis clusters:

- **my-cluster1** – a Redis (cluster mode disabled) cluster with 1 node.
- **my-cluster2** – a Redis (cluster mode disabled) cluster with 2 nodes, a primary and 1 read replica.
- **my-cluster22** – a Redis (cluster mode enabled) cluster with 2 shards, each shard having a primary node and 2 read replicas. This cluster has encryption enabled.

```json
{
    "CacheClusters": [
        {
            "CacheClusterStatus": "available",
            "SecurityGroups": [
                {
                    "Status": "active",
                    "SecurityGroupId": "sg-dbe93fa2"
                }
            ],
            "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
            "Engine": "redis",
            "PreferredMaintenanceWindow": "wed:12:00-wed:13:00",
            "CacheSubnetGroupName": "default",
            "SnapshotWindow": "08:30-09:30",
            "TransitEncryptionEnabled": false,
            "AtRestEncryptionEnabled": false,
            "CacheClusterId": "my-cluster1",
            "CacheClusterCreateTime": "2018-02-26T21:06:43.420Z",
            "PreferredAvailabilityZone": "us-west-2c",
            "AuthTokenEnabled": false,
            "PendingModifiedValues": {},
            "CacheNodeType": "cache.r4.large",
            "CacheParameterGroup": {
                "CacheNodeIdsToReboot": [],
                "ParameterApplyStatus": "in-sync",
                "CacheParameterGroupName": "default.redis3.2"
            },
            "SnapshotRetentionLimit": 0,
            "AutoMinorVersionUpgrade": true,
            "EngineVersion": "3.2.10",
            "CacheSecurityGroups": [],
            "NumCacheNodes": 1
        },
        {
            "CacheClusters": [
                {
                    "SecurityGroups": [
                        {
                            "Status": "active",
                            "SecurityGroupId": "sg-dbe93fa2"
                        }
                    ],
                    "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
                    "Engine": "redis",
                    "PreferredMaintenanceWindow": "wed:12:00-wed:13:00",
                    "CacheSubnetGroupName": "default",
                    "SnapshotWindow": "08:30-09:30",
                    "TransitEncryptionEnabled": false,
                    "AtRestEncryptionEnabled": false,
                    "CacheClusterId": "my-cluster2",
                    "CacheClusterCreateTime": "2018-02-26T21:06:43.420Z",
                    "PreferredAvailabilityZone": "us-west-2c",
                    "AuthTokenEnabled": false,
                    "PendingModifiedValues": {},
                    "CacheNodeType": "cache.r4.large",
                    "CacheParameterGroup": {
                        "CacheNodeIdsToReboot": [],
                        "ParameterApplyStatus": "in-sync",
                        "CacheParameterGroupName": "default.redis3.2"
                    },
                    "SnapshotRetentionLimit": 0,
                    "AutoMinorVersionUpgrade": true,
                    "EngineVersion": "3.2.10",
                    "CacheSecurityGroups": [],
                    "NumCacheNodes": 2
                },
                {
                    "SecurityGroups": [
                        {
                            "Status": "active",
                            "SecurityGroupId": "sg-dbe93fa2"
                        }
                    ],
                    "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
                    "Engine": "redis",
                    "PreferredMaintenanceWindow": "wed:12:00-wed:13:00",
                    "CacheSubnetGroupName": "default",
                    "SnapshotWindow": "08:30-09:30",
                    "TransitEncryptionEnabled": false,
                    "AtRestEncryptionEnabled": false,
                    "CacheClusterId": "my-cluster22",
                    "CacheClusterCreateTime": "2018-02-26T21:06:43.420Z",
                    "PreferredAvailabilityZone": "us-west-2c",
                    "AuthTokenEnabled": false,
                    "PendingModifiedValues": {},
                    "CacheNodeType": "cache.r4.large",
                    "CacheParameterGroup": {
                        "CacheNodeIdsToReboot": [],
                        "ParameterApplyStatus": "in-sync",
                        "CacheParameterGroupName": "default.redis3.2"
                    },
                    "SnapshotRetentionLimit": 0,
                    "AutoMinorVersionUpgrade": true,
                    "EngineVersion": "3.2.10",
                    "CacheSecurityGroups": [],
                    "NumCacheNodes": 2
                }
            ]
        }
    ]
}
```
"AuthTokenEnabled": false,
"CacheSubnetGroupName": "default",
"SnapshotWindow": "12:30-13:30",
"AutoMinorVersionUpgrade": true,
"CacheClusterStatus": "available",
"AtRestEncryptionEnabled": false,
"PreferredAvailabilityZone": "us-west-2a",
"TransitEncryptionEnabled": false,
"ReplicationGroupId": "my-cluster2",
"Engine": "redis",
"PreferredMaintenanceWindow": "sun:08:30-sun:09:30",
"CacheClusterId": "my-cluster2-001",
"PendingModifiedValues": {},
"CacheNodeType": "cache.r4.large",
"CacheParameterGroup": {
  "CacheNodeIdsToReboot": [],
  "ParameterApplyStatus": "in-sync",
  "CacheParameterGroupName": "default.redis3.2"
},
"SnapshotRetentionLimit": 0,
"EngineVersion": "3.2.10",
"CacheSecurityGroups": [],
"NumCacheNodes": 1
},
{
  "SecurityGroups": [
    {
      "Status": "active",
      "SecurityGroupId": "sg-dbe93fa2"
    }
  ],
  "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:"
}
Amazon ElastiCache for Redis
ElastiCache for Redis User Guide
Viewing a Cluster's Details

```json
{
  "CacheClusters": [
    {
      "SecurityGroups": [
        {
          "Status": "active",
          "SecurityGroupId": "sg-dbe93fa2"
        }
      ],
      "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
      "AuthTokenEnabled": true,
      "CacheSubnetGroupName": "default",
      "SnapshotWindow": "12:30-13:30",
      "AutoMinorVersionUpgrade": true,
      "CacheClusterCreateTime": "2018-02-26T21:17:01.439Z",
      "CacheClusterStatus": "available",
      "AtRestEncryptionEnabled": true,
      "PreferredAvailabilityZone": "us-west-2a",
      "TransitEncryptionEnabled": true,
      "ReplicationGroupId": "my-cluster3",
      "Engine": "redis",
      "PreferredMaintenanceWindow": "thu:11:00-thu:12:00",
      "CacheClusterId": "my-cluster3-0001-001",
      "PendingModifiedValues": {},
      "CacheNodeType": "cache.r4.large",
      "CacheParameterGroup": {
        "CacheNodeIdsToReboot": [],
        "ParameterApplyStatus": "in-sync",
        "CacheParameterGroupName": "default.redis3.2.cluster.on"
      },
      "SnapshotRetentionLimit": 0,
      "EngineVersion": "3.2.6",
      "CacheSecurityGroups": [],
      "NumCacheNodes": 1
    },
    {
      "SecurityGroups": [
        {
          "Status": "active",
          "SecurityGroupId": "sg-dbe93fa2"
        }
      ],
      "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
      "AuthTokenEnabled": true,
      "CacheSubnetGroupName": "default",
      "SnapshotWindow": "12:30-13:30",
      "AutoMinorVersionUpgrade": true,
      "CacheClusterStatus": "available",
      "AtRestEncryptionEnabled": false,
      "PreferredAvailabilityZone": "us-west-2c",
      "TransitEncryptionEnabled": false,
      "ReplicationGroupId": "my-cluster2",
      "Engine": "redis",
      "PreferredMaintenanceWindow": "sun:08:30-sun:09:30",
      "CacheClusterId": "my-cluster2-003",
      "PendingModifiedValues": {},
      "CacheNodeType": "cache.r4.large",
      "CacheParameterGroup": {
        "CacheNodeIdsToReboot": [],
        "ParameterApplyStatus": "in-sync",
        "CacheParameterGroupName": "default.redis3.2"
      },
      "SnapshotRetentionLimit": 0,
      "EngineVersion": "3.2.10",
      "CacheSecurityGroups": [],
      "NumCacheNodes": 1
    }
  ]
}
```


```json
{
    "SecurityGroups": [ 
    {
        "Status": "active",
        "SecurityGroupId": "sg-dbe93fa2"
    }],
    "AuthTokenEnabled": true,
    "CacheSubnetGroupName": "default",
    "AutoMinorVersionUpgrade": true,
    "CacheClusterCreateTime": "2018-02-26T21:17:01.439Z",
    "CacheClusterStatus": "available",
    "AtRestEncryptionEnabled": true,
    "PreferredAvailabilityZone": "us-west-2b",
    "TransitEncryptionEnabled": true,
    "ReplicationGroupId": "my-cluster3",
    "Engine": "redis",
    "PreferredMaintenanceWindow": "thu:11:00-thu:12:00",
    "CacheClusterId": "my-cluster3-0001-002",
    "PendingModifiedValues": {},
    "CacheNodeType": "cache.r4.large",
    "CacheParameterGroup": {
        "CacheNodeIdsToReboot": [],
        "ParameterApplyStatus": "in-sync",
        "CacheParameterGroupName": "default.redis3.2.cluster.on"
    },
    "SnapshotRetentionLimit": 0,
    "EngineVersion": "3.2.6",
    "CacheSecurityGroups": [],
    "NumCacheNodes": 1
},
{
    "SecurityGroups": [ 
    {
        "Status": "active",
        "SecurityGroupId": "sg-dbe93fa2"
    }],
    "AuthTokenEnabled": true,
    "CacheSubnetGroupName": "default",
    "AutoMinorVersionUpgrade": true,
    "CacheClusterCreateTime": "2018-02-26T21:17:01.439Z",
    "CacheClusterStatus": "available",
    "AtRestEncryptionEnabled": true,
    "PreferredAvailabilityZone": "us-west-2c",
    "TransitEncryptionEnabled": true,
    "ReplicationGroupId": "my-cluster3",
    "Engine": "redis",
    "PreferredMaintenanceWindow": "thu:11:00-thu:12:00",
    "CacheClusterId": "my-cluster3-0001-003",
    "PendingModifiedValues": {},
    "CacheNodeType": "cache.r4.large",
    "CacheParameterGroup": {
        "CacheNodeIdsToReboot": [],
        "ParameterApplyStatus": "in-sync",
        "CacheParameterGroupName": "default.redis3.2.cluster.on"
    },
    "SnapshotRetentionLimit": 0,
    "EngineVersion": "3.2.6",
    "CacheSecurityGroups": []
}
```
"NumCacheNodes": 1
},
"SecurityGroups": [
{
"Status": "active",
"SecurityGroupId": "sg-dbe93fa2"
}
],
"ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
"AuthTokenEnabled": true,
"CacheSubnetGroupName": "default",
"SnapshotWindow": "12:30-13:30",
"AutoMinorVersionUpgrade": true,
"CacheClusterCreateTime": "2018-02-26T21:17:01.439Z",
"CacheClusterStatus": "available",
"AtRestEncryptionEnabled": true,
"PreferredAvailabilityZone": "us-west-2b",
"TransitEncryptionEnabled": true,
"ReplicationGroupId": "my-cluster3",
"Engine": "redis",
"PreferredMaintenanceWindow": "thu:11:00-thu:12:00",
"CacheClusterId": "my-cluster3-0002-001",
"PendingModifiedValues": {},
"CacheNodeType": "cache.r4.large",
"CacheParameterGroup": {
"CacheNodeIdsToReboot": [],
"ParameterApplyStatus": "in-sync",
"CacheParameterGroupName": "default.redis3.2.cluster.on"
},
"SnapshotRetentionLimit": 0,
"EngineVersion": "3.2.6",
"CacheSecurityGroups": [],
"NumCacheNodes": 1
},
"SecurityGroups": [
{
"Status": "active",
"SecurityGroupId": "sg-dbe93fa2"
}
],
"ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
"AuthTokenEnabled": true,
"CacheSubnetGroupName": "default",
"SnapshotWindow": "12:30-13:30",
"AutoMinorVersionUpgrade": true,
"CacheClusterCreateTime": "2018-02-26T21:17:01.439Z",
"CacheClusterStatus": "available",
"AtRestEncryptionEnabled": true,
"PreferredAvailabilityZone": "us-west-2c",
"TransitEncryptionEnabled": true,
"ReplicationGroupId": "my-cluster3",
"Engine": "redis",
"PreferredMaintenanceWindow": "thu:11:00-thu:12:00",
"CacheClusterId": "my-cluster3-0002-002",
"PendingModifiedValues": {},
"CacheNodeType": "cache.r4.large",
"CacheParameterGroup": {
"CacheNodeIdsToReboot": [],
"ParameterApplyStatus": "in-sync",
"CacheParameterGroupName": "default.redis3.2.cluster.on"
},
"SnapshotRetentionLimit": 0,
"EngineVersion": "3.2.6",
"CacheSecurityGroups": []
"NumCacheNodes": 1
},

"SecurityGroups": [
{
    "Status": "active",
    "SecurityGroupId": "sg-dbe93fa2"
}
],


"AuthTokenEnabled": true,
"CacheSubnetGroupName": "default",
"SnapshotWindow": "12:30-13:30",
"AutoMinorVersionUpgrade": true,
"CacheClusterCreateTime": "2018-02-26T21:17:01.439Z",
"CacheClusterStatus": "available",
"AtRestEncryptionEnabled": true,
"PreferredAvailabilityZone": "us-west-2a",
"TransitEncryptionEnabled": true,
"ReplicationGroupId": "my-cluster3",
"Engine": "redis",
"PreferredMaintenanceWindow": "thu:11:00-thu:12:00",
"CacheClusterId": "my-cluster3-0002-003",
"PendingModifiedValues": {},
"CacheNodeType": "cache.r4.large",
"CacheParameterGroup": {
    "CacheNodeIdsToReboot": [],
    "ParameterApplyStatus": "in-sync",
    "CacheParameterGroupName": "default.redis3.2.cluster.on"
},
"SnapshotRetentionLimit": 0,
"EngineVersion": "3.2.6",
"CacheSecurityGroups": [],
"NumCacheNodes": 1
}

For more information, see the AWS CLI for ElastiCache topic describe-cache-clusters.

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
&X-Amz-Credential=<credential>

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

API Version 2015-02-02
98
For more information, see the ElastiCache API reference topic `DescribeCacheClusters`.

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>
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 Redis Specific Parameters (p. 315). For information on rebooting a cluster, see Rebooting a Cluster (p. 103).

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. 53). However, you can't downgrade to older engine versions except by deleting the existing cluster and creating it again.

Because the newer Redis versions provide a better and more stable user experience, Redis versions 2.6.13, 2.8.6, and 2.8.19 are deprecated when using the ElastiCache console. We recommend against using these Redis versions. If you need to use one of them, work with the AWS CLI or ElastiCache API.

For more information, see the following topics:

<table>
<thead>
<tr>
<th>Action</th>
<th>AWS CLI</th>
<th>ElastiCache API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Cluster</td>
<td>Creating a Cluster (AWS CLI) (p. 87)</td>
<td>Creating a Cluster (ElastiCache API) (p. 88)</td>
</tr>
<tr>
<td></td>
<td>You can't use this action to create a replication group with</td>
<td>You can't use this action to create a replication group with</td>
</tr>
<tr>
<td></td>
<td>cluster mode enabled.</td>
<td>cluster mode enabled.</td>
</tr>
<tr>
<td>Modify Cluster</td>
<td>Using the AWS CLI (p. 101)</td>
<td>Using the ElastiCache API (p. 102)</td>
</tr>
</tbody>
</table>
### AWS CLI

<table>
<thead>
<tr>
<th>Create Replication Group</th>
<th>You can't use this action to create a replication group with cluster mode enabled.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElastiCache API</td>
<td>You can't use this action to create a replication group with cluster mode enabled.</td>
</tr>
</tbody>
</table>

#### Create Replication Group

- **Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch (AWS CLI)** (p. 169)
- **Creating a Redis (Cluster Mode Enabled) Replication Group from Scratch (AWS CLI)** (p. 175)

#### Modify Replication Group

- **Using the AWS CLI** (p. 188)
- **Using the ElastiCache API** (p. 189)

---

The **Apply Immediately** box applies only to node type and engine version modifications. To apply changes immediately, choose the **Apply Immediately** check box. If this box is not chosen, node type and 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. 53). However, you can't downgrade to older engine versions except by deleting the existing cluster or replication group and creating it again.

For Linux, macOS, or Unix:

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

For Windows:

```bash
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`.

---

**API Version 2015-02-02**

**101**
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. 53). However, you can't downgrade to older engine versions except by deleting the existing cluster or replication group and creating it again.

Line breaks are added for ease of reading.

```
https://elasticache.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`.
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.

Redis (Cluster Mode Enabled) and Reboots
If you make changes to parameters that require a Redis (cluster mode enabled) cluster reboot for the changes to be applied, follow these steps.

2. Delete the Redis (cluster mode enabled) cluster. See Deleting a Cluster (p. 117).
3. Restore the cluster using the altered parameter group and backup to seed the new cluster. See Restoring From a Backup with Optional Cluster Resizing (p. 239).

Using the AWS Management Console

You can reboott 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:

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

For Windows:

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

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`.

```
```

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

To reconfigure your Redis (cluster mode enabled) cluster, see Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)

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

### Using the AWS Management Console

If you want to add a node to a single-node Redis (cluster mode disabled) cluster (one without replication enabled), it's a two-step process: first add replication, and then add a replica node.

#### Topics

- To add replication to a Redis cluster with no shards (p. 105)
- To add nodes to a cluster (console) (p. 105)

The following procedure adds replication to a single-node Redis that does not have replication enabled. When you add replication, the existing node becomes the primary node in the replication-enabled cluster. After replication is added, you can add up to 5 replica nodes to the cluster.
To add replication to a Redis cluster with no shards

2. From the navigation pane, choose Redis.

   A list of clusters running the Redis engine is displayed.
3. Choose the name of a cluster, not the box to the left of the cluster's name, that you want to add nodes to.

   The following is true of a Redis cluster that does not have replication enabled:
   • It is running Redis, not Clustered Redis.
   • It has zero shards.

   If the cluster has any shards, replication is already enabled on it and you can continue at To add nodes to a cluster (console) (p. 105).
4. Choose Add replication.
5. In Add Replication, type a description for this replication-enabled cluster.
6. Choose Add.

   As soon as the cluster's status returns to available you can continue at the next procedure and add replicas to the cluster.

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.

   If your cluster is a Redis (cluster mode enabled) cluster, see Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279).

   If your cluster is a Redis (cluster mode disabled) cluster with zero shards, first complete the steps at To add replication to a Redis cluster with no shards (p. 105).
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</td>
</tr>
</tbody>
</table>
### Scenarios

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Pending Operation</th>
<th>New Request</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<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. <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.
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

If you want to add nodes to an existing Redis (cluster mode disabled) cluster that does not have replication enabled, you must first create the replication group specifying the existing cluster as the primary. For more information, see Creating a Replication Group Using an Available Redis Cache Cluster (AWS CLI) (p. 163). After the replication group is **available**, you can continue with the following process.

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. 111).
- **--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:

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

For Windows:

API Version 2015-02-02
aws elasticache modify-cache-cluster
  --cache-cluster-id my-cluster
  --num-cache-nodes 5
  --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-west-2b",
    "ConfigurationEndpoint": {
      "Port": 11211,
      "Address": "rlh-mem000.7alc7bf-example.cfg.usw2.cache.amazonaws.com"
    },
    "CacheSecurityGroups": [],
    "AutoMinorVersionUpgrade": true,
    "CacheClusterStatus": "modifying",
    "NumCacheNodes": 2,
    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

If you want to add nodes to an existing Redis (cluster mode disabled) cluster that does not have replication enabled, you must first create the replication group specifying the existing cluster as the Primary. For more information, see Adding Replicas to a Standalone Redis (Cluster Mode Disabled) Cluster (ElastiCache API) (p. 165). After the replication group is available, you can continue with the following process.

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. 114)).

- **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**

```plaintext
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**.
Removing Nodes from a Cluster

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. 110)
- Using the AWS CLI (p. 111)
- Using the ElastiCache API (p. 114)

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, 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.</td>
</tr>
</tbody>
</table>
### Scenarios

<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>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 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.</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. 89).
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:

```bash
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:

```bash
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):

```json
{
  "CacheClusters": [
    {
      "SecurityGroups": [
        {
          "Status": "active",
          "SecurityGroupId": "sg-dbe93fa2"
        }
      ],
      "ClientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download:",
      "AuthTokenEnabled": false,
      "CacheSubnetGroupName": "default",
      "SnapshotWindow": "12:30-13:30",
      "AutoMinorVersionUpgrade": true,
      "CacheClusterStatus": "available",
      "AtRestEncryptionEnabled": false,
      "PreferredAvailabilityZone": "us-west-2a",
      "TransitEncryptionEnabled": false,
      "ReplicationGroupId": "my-cluster2",
      "Engine": "redis",
      "PreferredMaintenanceWindow": "sun:08:30-sun:09:30",
      "CacheClusterId": "my-cluster2-001",
      "PendingModifiedValues": {},
      "API Version 2015-02-02"
    }
  ]
}
```
Removing Nodes from a Cluster

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.

API Version 2015-02-02
114
Removing Nodes from a 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`.

```text
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
&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`.
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

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.

```bash
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:

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

For Windows:

```bash
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`. 
Accessing Your Cluster or Replication Group

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. 119)
• Grant Access to Your Cluster or Replication Group (p. 121)

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 Redis engine, in the left navigation pane, choose Redis.
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. 121).
   • 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. 121).

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": "redis",
         "AuthTokenEnabled": false,
         "CacheParameterGroup": {
           "CacheNodeIdsToReboot": [],
           "CacheParameterGroupName": "default.redis3.2",
           "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",
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"
       }
     ]
   }
   ```
Grant Access to Your Cluster or Replication Group

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 Redis is 6379.
   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. 121).

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. 382).
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. 123)
- Configuring AWS Client VPN Components (p. 124)
- Configure the VPN Client (p. 126)

**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-rsa`
  - `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:

```bash
```

Example output:

```
"ClientVpnEndpointId": "cvpn-endpoint-0123456789abcdefg",
"Status": { "Code": "pending-associate" }, "DnsName": "cvpn-endpoint-0123456789abcdefg.prod.clientvpn.us-east-1.amazonaws.com"
```

### Example output:

```
"ClientVpnEndpointId": "cvpn-endpoint-0123456789abcdefg",
"Status": { "Code": "pending-associate" }, "DnsName": "cvpn-endpoint-0123456789abcdefg.prod.clientvpn.us-east-1.amazonaws.com"
```
Run the following command:

```bash
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:

```bash
aws ec2 apply-security-groups-to-client-vpn-target-network --client-vpn-endpoint-id cvpn-endpoint-0123456789abcdefga --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:

```bash
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:

```
API Version 2015-02-02
125
```
Grant Access to Your Cluster or Replication Group

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.

**Using the AWS CLI**

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 }
```

## Working with Shards

A shard (API/CLI: node group) is a collection of one to six Redis nodes. A Redis (cluster mode disabled) cluster will never have more than one shard. Redis (cluster mode enabled) clusters can have from 1 to 90 shards. You can create a cluster with higher number of shards and lower number of replicas totaling up to 90 nodes per cluster. This cluster configuration can range from 90 shards and 0 replicas to 15 shards and 5 replicas, which is the maximum number of replicas allowed. The cluster's data is partitioned across the cluster's shards. If there is more than one node in a shard, the shard implements replication with one node being the read/write primary node and the other nodes read-only replica nodes.

**Note**

The node or shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see [AWS Service Limits](https://docs.aws.amazon.com/service-quotas/latest/guides/quotas.html) and select limit type "Nodes per cluster per instance type".

When you create a Redis (cluster mode enabled) cluster using the ElastiCache console, you specify the number of shards in the cluster and the number of nodes in the shards. For more information, see [Creating a Redis (Cluster Mode Enabled) Cluster (Console)](https://docs.aws.amazon.com/elasticsearch-service/latest/redis Deletes/creating-a-redis-cluster-mode-enabled-cluster.html). If you use the ElastiCache API or AWS CLI to create a cluster (called replication group in the API/CLI), you can configure the number of nodes in a shard (API/CLI: node group) independently. For more information, see the following:

- **API**: CreateReplicationGroup
- **CLI**: create-replication-group

Each node in a shard has the same compute, storage and memory specifications. The ElastiCache API lets you control shard-wide attributes, such as the number of nodes, security settings, and system maintenance windows.
Redis shard configurations

Finding a Shard's ID

You can find a shard's ID using the AWS Management Console, the AWS CLI or the ElastiCache API.

Topics
- Using the AWS Management Console (p. 128)
- Using the AWS CLI (p. 128)
- Using the ElastiCache API (p. 130)

Using the AWS Management Console

Topics
- For Redis (Cluster Mode Disabled) (p. 128)
- For Redis (Cluster Mode Enabled) (p. 128)

For Redis (Cluster Mode Disabled)

Redis (cluster mode disabled) replication group shard IDs are always 0001.

For Redis (Cluster Mode Enabled)

The following procedure uses the AWS Management Console to find a Redis (cluster mode enabled)'s replication group's shard ID.

To find the shard ID in a Redis (cluster mode enabled) replication group

2. On the navigation pane, choose Redis, then choose the name of the Redis (cluster mode enabled) replication group you want to find the shard IDs for.
3. In the Shard Name column, the shard ID is the last four digits of the shard name.

Using the AWS CLI

To find shard (node group) ids for either Redis (cluster mode disabled) or Redis (cluster mode enabled) replication groups use the AWS CLI operation describe-replication-groups with the following optional parameter.

- --replication-group-id—An optional parameter which when used limits the output to the details of the specified replication group. If this parameter is omitted, the details of up to 100 replication groups is returned.

Example

This command returns the details for sample-repl-group.

For Linux, macOS, or Unix:

```
aws elasticache describe-replication-groups \
  --replication-group-id sample-repl-group
```
For Windows:

```bash
aws elasticache describe-replication-groups --replication-group-id sample-repl-group
```

Output from this command looks something like this. The shard (node group) ids are highlighted here to make finding them easier.

```
{
    "ReplicationGroups": [
    {
        "Status": "available",
        "Description": "2 shards, 2 nodes (1 + 1 replica)",
        "Nodes": [
            {
                "Status": "available",
                "Slots": "0-8191",
                "NodeId": "0001",
                "NodeGroupMembers": [
                {
                    "PreferredAvailabilityZone": "us-west-2c",
                    "CacheNodeId": "0001",
                    "CacheClusterId": "sample-repl-group-0001-001"
                },
                {
                    "PreferredAvailabilityZone": "us-west-2a",
                    "CacheNodeId": "0001",
                    "CacheClusterId": "sample-repl-group-0001-002"
                }
            ]
        }
    },
    {
        "Status": "available",
        "Slots": "8192-16383",
        "NodeId": "0002",
        "NodeGroupMembers": [
            {
                "PreferredAvailabilityZone": "us-west-2b",
                "CacheNodeId": "0001",
                "CacheClusterId": "sample-repl-group-0002-001"
            },
            {
                "PreferredAvailabilityZone": "us-west-2a",
                "CacheNodeId": "0001",
                "CacheClusterId": "sample-repl-group-0002-002"
            }
        ]
    }
    ],
    "ConfigurationEndpoint": {
        "Port": 6379,
        "Address": "sample-repl-group.9dcv5r.clustercfg.usw2.cache.amazonaws.com"
    },
    "ClusterEnabled": true,
    "ReplicationGroupId": "sample-repl-group",
    "SnapshotRetentionLimit": 1,
    "AutomaticFailover": "enabled",
    "SnapshotWindow": "13:00-14:00",
    "MemberClusters": [
        "sample-repl-group-0001-001",
        "sample-repl-group-0001-002",
        "sample-repl-group-0002-001",
        "sample-repl-group-0002-002"
    ]
}
```
Using the ElastiCache API

To find shard (node group) ids for either Redis (cluster mode disabled) or Redis (cluster mode enabled) replication groups use the AWS CLI operation `describe-replication-groups` with the following optional parameter.

- **ReplicationGroupId**—An optional parameter which when used limits the output to the details of the specified replication group. If this parameter is omitted, the details of up to \( \text{xxx} \) replication groups is returned.

Example

This command returns the details for `sample-repl-group`.

For Linux, macOS, or Unix:

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeReplicationGroup
&ReplicationGroupId=sample-repl-group
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>
```

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. 130)
- Prerequisites and Limitations (p. 132)
- Using Global Datastores (Console) (p. 132)
- Using Global Datastores (CLI) (p. 140)

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.

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. 363). 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. 139).
- 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. 133)
- Creating a New Global Datastore Using a New Primary Cluster (p. 134)
- Viewing Global Datastore Details (p. 136)
- Adding a Region to a Global Datastore (p. 137)
- Modifying a Global Datastore (p. 138)
- Promoting the Secondary Cluster to Primary (p. 139)
- Removing a Region from a Global Datastore (p. 139)
- Deleting a Global Datastore (p. 140)

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. 137).

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

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. 135).
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.

   - 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
   - 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

API Version 2015-02-02
• 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. 137).

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. 133) 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. 137)
• Removing a Region from a Global Datastore (p. 139)
• Promoting the Secondary Cluster to Primary (p. 139)
• Modifying a Global Datastore (p. 138)

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
• Redis AUTH

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. 139).

**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. 65).
• **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. 310). 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. 139). 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
```

- **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 \
  --global-replication-group-id my global datastore
```
High Availability Using Replication Groups

Single-node Amazon ElastiCache Redis clusters are in-memory entities with limited data protection services (AOF). If your cluster fails for any reason, you lose all the cluster's data. However, if you're running the Redis engine, you can group 2 to 6 nodes into a cluster with replicas where 1 to 5 read-only nodes contain replicate data of the group's single read/write primary node. In this scenario, if one node fails for any reason, you do not lose all your data since it is replicated in one or more other nodes. Due to replication latency, some data may be lost if it is the primary read/write node that fails.

As seen in the following graphic, the replication structure is contained within a shard (called node group in the API/CLI) which is contained within a Redis cluster. Redis (cluster mode disabled) clusters always have one shard. Redis (cluster mode enabled) clusters can have up to 90 shards with the cluster's data partitioned across the shards. You can create a cluster with higher number of shards and lower number of replicas totaling up to 90 nodes per cluster. This cluster configuration can range from 90 shards and 0 replicas to 15 shards and 5 replicas, which is the maximum number of replicas allowed.

**Note**
The node or shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see AWS Service Limits and select limit type "Nodes per cluster per instance type".

---

**High Availability Using Replication Groups**

**Syntax**

```bash
aws elasticache increase-node-groups-in-global-replication-group
    --apply-immediately yes
    --global-replication-group-id global-replication-group-name
    --node-group-count 3
```

```bash
aws elasticache decrease-node-groups-in-global-replication-group
    --apply-immediately yes
    --global-replication-group-id global-replication-group-name
    --node-group-count 3
```

```bash
aws elasticache rebalance-shards-in-global-replication-group
    --apply-immediately yes
    --global-replication-group-id global-replication-group-name
```

**Use help to list all available commands ElastiCache for Redis.**

```bash
aws elasticache help
```

**You can also use help to describe a specific command and learn more about its usage:**

```bash
aws elasticache create-global-replication-group help
```
Redis (cluster mode disabled) cluster has one shard and 0 to 5 replica nodes

If the cluster with replicas has Multi-AZ with Automatic Failover enabled and the primary node fails, the primary fails over to a read replica. Because the data is updated on the replica nodes asynchronously, there may be some data loss due to latency in updating the replica nodes. For more information, see Mitigating Failures when Running Redis (p. 454).

Topics

- Understanding Redis Replication (p. 144)
- Replication: Redis (Cluster Mode Disabled) vs. Redis (Cluster Mode Enabled) (p. 145)
- Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148)
- How Synchronization and Backup are Implemented (p. 161)
- Creating a Redis Replication Group (p. 162)
- Viewing a Replication Group's Details (p. 182)
- Finding Replication Group Endpoints (p. 187)
- Modifying a Replication Group (p. 188)
- Deleting a Replication Group (p. 190)
- Changing the Number of Replicas (p. 191)
- Promoting a Read Replica to Primary, for Redis (cluster mode disabled) Replication Groups (p. 203)
Understanding Redis Replication

Redis implements replication in two ways:

- With a single shard that contains all of the cluster's data in each node—Redis (cluster mode disabled)
- With data partitioned across up to 90 shards—Redis (cluster mode enabled)

Each shard in a replication group has a single read/write primary node and up to 5 read-only replica nodes. You can create a cluster with higher number of shards and lower number of replicas totaling up to 90 nodes per cluster. This cluster configuration can range from 90 shards and 0 replicas to 15 shards and 5 replicas, which is the maximum number of replicas allowed.

**Note**
The node or shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see AWS Service Limits and select limit type "Nodes per cluster per instance type".

**Topics**
- Redis (Cluster Mode Disabled) (p. 144)
- Redis (cluster mode enabled) (p. 145)

Redis (Cluster Mode Disabled)

A Redis (cluster mode disabled) cluster has a single shard, inside of which is a collection of Redis nodes; one primary read/write node and up to five secondary, read-only replica nodes. Each read replica maintains a copy of the data from the cluster's primary node. Asynchronous replication mechanisms are used to keep the read replicas synchronized with the primary. Applications can read from any node in the cluster. Applications can write only to the primary node. Read replicas improve read throughput and guard against data loss in cases of a node failure.

You can use Redis (cluster mode disabled) clusters with replica nodes to scale your Redis solution for ElastiCache to handle applications that are read-intensive or to support large numbers of clients that simultaneously read from the same cluster.

All of the nodes in a Redis (cluster mode disabled) cluster must reside in the same region. To improve fault tolerance, you can provision read replicas in multiple Availability Zones within that region.

When you add a read replica to a cluster, all of the data from the primary is copied to the new node. From that point on, whenever data is written to the primary, the changes are asynchronously propagated to all the read replicas.

To improve fault tolerance and reduce write downtime, enable Multi-AZ with Automatic Failover for your Redis (cluster mode disabled) cluster with replicas. For more information, see Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148).

You can change the roles of the nodes within the Redis (cluster mode disabled) cluster, with the primary and one of the replicas exchanging roles. You might decide to do this for performance tuning reasons. For example, with a web application that has heavy write activity, you can choose the node that has
Replication: Redis (Cluster Mode Disabled) vs. Redis (Cluster Mode Enabled)

Beginning with Redis version 3.2, you have the ability to create one of two distinct types of Redis clusters (API/CLI: replication groups). A Redis (cluster mode disabled) cluster always has a single shard (API/CLI: node group). Each shard has one primary node and up to five read-only replica nodes. The configuration can range from 90 shards and 0 replicas to 15 shards and 5 replicas, which is the maximum number or replicas allowed. Each read replica in a shard maintains a copy of the data from the shard's primary. Asynchronous replication mechanisms are used to keep the read replicas synchronized with the primary. Applications can read from any node in the cluster. Applications can write only to the primary nodes. Read replicas enhance read scalability and guard against data loss. Data is partitioned across the shards in a Redis (cluster mode enabled) cluster.

**Note**

The node/shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see AWS Service Limits and include the instance type in the request.

Applications use the Redis (cluster mode enabled) cluster’s *configuration endpoint* to connect with the nodes in the cluster. For more information, see Finding Connection Endpoints (p. 204).

**Redis (cluster mode enabled)**

A Redis (cluster mode enabled) cluster is comprised of from 1 to 90 shards (API/CLI: node groups). Each shard has a primary node and up to five read-only replica nodes. The configuration can range from 90 shards and 0 replicas to 15 shards and 5 replicas, which is the maximum number or replicas allowed. Each read replica in a shard maintains a copy of the data from the shard's primary. Asynchronous replication mechanisms are used to keep the read replicas synchronized with the primary. Applications can read from any node in the cluster. Applications can write only to the primary nodes. Read replicas enhance read scalability and guard against data loss. Data is partitioned across the shards in a Redis (cluster mode enabled) cluster.

---

**Redis (cluster mode enabled) cluster with multiple shards and replica nodes**

All of the nodes in a Redis (cluster mode enabled) cluster must reside in the same region. To improve fault tolerance, you can provision both primaries and read replicas in multiple Availability Zones within that region.

Multi-AZ with Automatic Failover is required for all Redis (cluster mode enabled) clusters. For more information, see Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148).

Currently, in Redis (cluster mode enabled), there are some limitations.

- You cannot manually promote any of the replica nodes to primary.
- Multi-AZ with Automatic Failover is required.
- You can only change the structure of a cluster, the node type, and the number of nodes by restoring from a backup. For more information, see Restoring From a Backup with Optional Cluster Resizing (p. 239). The number of shards in a Redis (cluster mode enabled) cluster can be changed dynamically, while the cluster continues to serve read and write requests. For more information, see Online Resharding and Shard Rebalancing for Redis (cluster mode enabled) (p. 281).
Replication: Redis (Cluster Mode Disabled) vs. Redis (Cluster Mode Enabled)

A Redis (cluster mode disabled) cluster has up to 5 read replica nodes. A Redis (cluster mode enabled) cluster has up to 90 shards with 1 to 5 read replica nodes in each.

**Redis (cluster mode disabled) and Redis (cluster mode enabled) clusters**

The following table summarizes important differences between Redis (cluster mode disabled) and Redis (cluster mode enabled) clusters.

### Comparing Redis (Cluster Mode Disabled) and Redis (Cluster Mode Enabled) Clusters

<table>
<thead>
<tr>
<th>Feature</th>
<th>Redis (cluster mode disabled)</th>
<th>Redis (cluster mode enabled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modifiable</td>
<td>Yes. Supports adding and deleting replica nodes, and scaling up node type.</td>
<td>Limited. For more information, see <a href="#">Upgrading Engine Versions</a> and <a href="##">Scaling Clusters in Redis (Cluster Mode Enabled)</a>.</td>
</tr>
<tr>
<td>Data Partitioning</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Shards</td>
<td>1</td>
<td>1 to 90</td>
</tr>
<tr>
<td>Read replicas</td>
<td>0 to 5</td>
<td>0 to 5 per shard.</td>
</tr>
<tr>
<td></td>
<td><strong>Important</strong> If you have no replicas and the node fails, you experience total data loss.</td>
<td><strong>Important</strong> If you have no replicas and a node fails, you experience loss of all data in that shard.</td>
</tr>
<tr>
<td>Snapshots (Backups)</td>
<td>Yes, creating a single .rdb file.</td>
<td>Yes, creating a unique .rdb file for each shard.</td>
</tr>
<tr>
<td>Restore</td>
<td>Yes, using a single .rdb file from a Redis (cluster mode disabled) cluster.</td>
<td>Yes, using .rdb files from either a Redis (cluster mode disabled) or a Redis (cluster mode enabled) cluster.</td>
</tr>
<tr>
<td>Supported by</td>
<td>All Redis versions</td>
<td>Redis 3.2 and following</td>
</tr>
</tbody>
</table>
When choosing between Redis (cluster mode disabled) or Redis (cluster mode enabled), consider the following factors:

- **Scaling v. partitioning** – Business needs change. You need to either provision for peak demand or scale as demand changes. Redis (cluster mode disabled) supports scaling. You can scale read capacity by adding or deleting replica nodes, or you can scale capacity by scaling up to a larger node type. Both of these operations take time. For more information, see Scaling Redis (Cluster Mode Disabled) Clusters with Replica Nodes (p. 264).

Redis (cluster mode enabled) supports partitioning your data across up to 90 node groups. You can dynamically change the number of shards as your business needs change. One advantage of partitioning is that you spread your load over a greater number of endpoints, which reduces access bottlenecks during peak demand. Additionally, you can accommodate a larger data set since the data can be spread across multiple servers. For information on scaling your partitions, see Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279).

- **Node size v. number of nodes** – Because a Redis (cluster mode disabled) cluster has only one shard, the node type must be large enough to accommodate all the cluster's data plus necessary overhead. On the other hand, because you can partition your data across several shards when using a Redis (cluster mode enabled) cluster, the node types can be smaller, though you need more of them. For more information, see Choosing Your Node Size (p. 76).

- **Reads v. writes** – If the primary load on your cluster is applications reading data, you can scale a Redis (cluster mode disabled) cluster by adding and deleting read replicas. However, there is a maximum of 5 read replicas. If the load on your cluster is write-heavy, you can benefit from the additional write endpoints of a Redis (cluster mode enabled) cluster with multiple shards.

Whichever type of cluster you choose to implement, be sure to choose a node type that is adequate for your current and future needs.
Minimizing Downtime: Multi-AZ with Automatic Failover

In certain cases, ElastiCache for Redis detects and replaces a primary node. These cases include certain types of planned maintenance and the unlikely event of a primary node or Availability Zone failure.

This replacement results in some downtime for the cluster. If you have Multi-AZ with automatic failover enabled on the cluster, the downtime is minimized. In this case, the role of primary node fails over to one of the read replicas. There’s no need to create and provision a new primary node. This failover and replica promotion ensure that you can resume writing to the new primary as soon as promotion is complete.

ElastiCache also propagates the Domain Name Service (DNS) name of the promoted replica. It does so because then if your application is writing to the primary endpoint, no endpoint change is required in your application. If you are reading from individual endpoints, you need to change the read endpoint of the replica promoted to primary to the new replica's endpoint.

In case of planned node replacements, initiated due to maintenance updates or self-service updates:

- For ElastiCache for Redis Cluster, the planned node replacements will now complete while the cluster serves incoming write requests.
- For Redis Cluster mode disabled clusters with Multi-AZ AutoFailover enabled and running on the 5.0.5 or above engine, the planned node replacements complete while the cluster serves incoming write requests.
- For Redis Cluster mode disabled clusters with Multi-AZ AutoFailover enabled and running on the 5.0.4 or below engine, you may notice a brief write interruption, of up to a few seconds, associated with DNS updates. This process is much faster than re-creating and provisioning a new primary, which is the process if you don't enable Multi-AZ with automatic failover.

You can enable Multi-AZ with Automatic Failover using the ElastiCache Management Console, the AWS CLI, or the ElastiCache API.

Enabling ElastiCache Multi-AZ with automatic failover on your Redis cluster (in the API and CLI, replication group) improves your fault tolerance. This is true particularly in cases where your cluster's read/write primary cluster becomes unreachable or fails for any reason. Multi-AZ with automatic failover is only supported on Redis clusters that support replication.

Topics
- Important Notes on Redis Multi-AZ with Automatic Failover (p. 148)
- Failure Scenarios with Multi-AZ and Automatic Failover Responses (p. 150)
- Enabling Multi-AZ with Automatic Failover (p. 154)
- Testing Multi-AZ with Automatic Failover (p. 157)

Important Notes on Redis Multi-AZ with Automatic Failover

The following points should be noted for Redis Multi-AZ with Automatic Failover:

- Multi-AZ with Automatic Failover is supported on Redis version 2.8.6 and later.
- Redis Multi-AZ with Automatic Failover is not supported on T1 node types.
- Redis replication is asynchronous. Therefore, when a primary cluster fails over to a replica, a small amount of data might be lost due to replication lag.

When choosing the replica to promote to primary, ElastiCache for Redis chooses the replica with the least replication lag. In other words, it chooses the replica that is most current. Doing so helps
minimize the amount of lost data. The replica with the least replication lag can be in the same or different Availability Zone from the failed primary node.

- When you manually promote read replicas to primary, you can only do so when Multi-AZ with Automatic Failover is disabled. To promote a read replica to primary, take the following steps:
  1. Disable Multi-AZ with Automatic Failover on the cluster.
  2. Promote the read replica to primary.
  3. Re-enable Multi-AZ with Automatic Failover.

You cannot disable Multi-AZ with Automatic Failover on Redis (cluster mode enabled) clusters. Therefore, you cannot manually promote a replica to primary on any Redis (cluster mode enabled) cluster.

- ElastiCache for Redis Multi-AZ with Automatic Failover and append-only file (AOF) are mutually exclusive. If you enable one, you cannot enable the other.

- A node's failure can be caused by the rare event of an entire Availability Zone failing. In this case, the replica replacing the failed primary is created only when the Availability Zone is back up. For example, consider a replication group with the primary in AZ-a and replicas in AZ-b and AZ-c. If the primary fails, the replica with the least replication lag is promoted to primary cluster. Then, ElastiCache creates a new replica in AZ-a (where the failed primary was located) only when AZ-a is back up and available.

- A customer-initiated reboot of a primary does not trigger automatic failover. Other reboots and failures do trigger automatic failover.

- Whenever the primary is rebooted, it is cleared of data when it comes back online. When the read replicas see the cleared primary cluster, they clear their copy of the data, which causes data loss.

- After a read replica has been promoted, the other replicas sync with the new primary. After the initial sync, the replicas' content is deleted and they sync the data from the new primary, causing a brief interruption during which the replicas are not accessible. This sync process also causes a temporary load increase on the primary while syncing with the replicas. This behavior is native to Redis and isn't unique to ElastiCache Multi-AZ. For details regarding this Redis behavior, see http://redis.io/topics/replication.

**Important**

- For Redis version 2.8.22 and later, external replicas are not permitted.

- For Redis versions before 2.8.22, we recommend that you do not connect an external Redis replica to an ElastiCache Redis cluster that is Multi-AZ with Automatic Failover enabled. This is an unsupported configuration that can create issues that prevent ElastiCache from properly performing failover and recovery. If you need to connect an external Redis replica to an ElastiCache cluster, make sure that Multi-AZ with Automatic Failover is disabled before you make the connection.
Failure Scenarios with Multi-AZ and Automatic Failover Responses

Before the introduction of Multi-AZ with Automatic Failover, ElastiCache detected and replaced a cluster's failed nodes by recreating and reprovisioning the failed node. By enabling Multi-AZ with Automatic Failover, a failed primary node fails over to the replica with the least replication lag. The selected replica is automatically promoted to primary, which is much faster than creating and reprovisioning a new primary node. This process usually takes just a few seconds until you can write to the cluster again.

When Multi-AZ with Automatic Failover is enabled, ElastiCache continually monitors the state of the primary node. If the primary node fails, one of the following actions is performed depending on the nature of the failure.

Failure Scenarios
- When Only the Primary Node Fails (p. 150)
- When the Primary Node and Some Read Replicas Fail (p. 151)
- When the Entire Cluster Fails (p. 152)

When Only the Primary Node Fails

If only the primary node fails, the read replica with the least replication lag is promoted to primary, and a replacement read replica is created and provisioned in the same Availability Zone as the failed primary.

1. The failed primary node is taken offline.
2. The read replica with the least replication lag is promoted to primary.
3. A replacement read replica is launched and provisioned.

Automatic Failover for a failed primary node

What ElastiCache Multi-AZ with Automatic Failover does when only the primary node fails is the following:

1. The failed primary node is taken offline.
2. The read replica with the least replication lag is promoted to primary.

Writes can resume as soon as the promotion process is complete, typically just a few seconds. If your application is writing to the primary endpoint, there is no need to change the endpoint for writes or reads as ElastiCache propagates the DNS name of the promoted replica.
3. A replacement read replica is launched and provisioned.
The replacement read replica is launched in the Availability Zone that the failed primary node was in so that the distribution of nodes is maintained.

4. The replicas sync with the new primary node.

After the new replica is available, be aware of these effects:

- **Primary endpoint** – You don't need to make any changes to your application, because the DNS name of the new primary node is propagated to the primary endpoint.
- **Read endpoint** – The reader endpoint is automatically updated to point to the new replica nodes.

For information about finding the endpoints of a cluster, see the following topics:

- Finding a Redis (Cluster Mode Disabled) Cluster's Endpoints (Console) (p. 206)
- Finding the Endpoints for Replication Groups (AWS CLI) (p. 211)
- Finding Endpoints for Replication Groups (ElastiCache API) (p. 213)

### When the Primary Node and Some Read Replicas Fail

If the primary and at least one read replica fails, the available replica with the least replication lag is promoted to primary cluster. New read replicas are also created and provisioned in the same Availability Zones as the failed nodes and replica that was promoted to primary.

What ElastiCache Multi-AZ does when the primary node and some read replicas fail is the following:

1. The failed primary node and failed read replicas are taken offline.
2. The available replica with the least replication lag is promoted to primary node.

   Writes can resume as soon as the promotion process is complete, typically just a few seconds. If your application is writing to the primary endpoint, there is no need to change the endpoint for writes, because ElastiCache propagates the DNS name of the promoted replica.
3. Replacement replicas are created and provisioned.

   The replacement replicas are created in the Availability Zones of the failed nodes so that the distribution of nodes is maintained.
4. All clusters sync with the new primary node.
You need to make the following changes to your application after the new nodes are available:

- **Primary endpoint** – Do not make any changes to your application because the DNS name of the new primary node is propagated to the primary endpoint.
- **Read endpoint** – The read endpoint will be automatically updated to point to the new replica nodes.

For information about finding the endpoints of a replication group, see the following topics:

- Finding a Redis (Cluster Mode Disabled) Cluster’s Endpoints (Console) (p. 206)
- Finding the Endpoints for Replication Groups (AWS CLI) (p. 211)
- Finding Endpoints for Replication Groups (ElastiCache API) (p. 213)

### When the Entire Cluster Fails

If everything fails, all the nodes are recreated and provisioned in the same Availability Zones as the original nodes.

In this scenario, all the data in the cluster is lost due to the failure of every node in the cluster. This occurrence is rare.

What ElastiCache Multi-AZ does when the entire cluster fails is the following:

1. The failed primary node and read replicas are taken offline.
2. A replacement primary node is created and provisioned.
3. Replacement replicas are created and provisioned.

The replacements are created in the Availability Zones of the failed nodes so that the distribution of nodes is maintained.

Because the entire cluster failed, data is lost and all the new nodes start cold.

Because each of the replacement nodes will have the same endpoint as the node it is replacing, you don't need to make any endpoint changes in your application.

For information about finding the endpoints of a replication group, see the following topics:

- Finding a Redis (Cluster Mode Disabled) Cluster’s Endpoints (Console) (p. 206)
We recommend that you create the primary node and read replicas in different Availability Zones to raise your fault tolerance level.
Enabling Multi-AZ with Automatic Failover

You can enable Multi-AZ with Automatic Failover when you create or modify a cluster (API or CLI, replication group) using the ElastiCache console, AWS CLI, or the ElastiCache API.

You can enable Multi-AZ with Automatic Failover only on Redis (cluster mode disabled) clusters that have at least one available read replica. Multi-AZ with Automatic Failover is required on all Redis (cluster mode enabled) clusters, whether or not they have read replicas. Clusters without read replicas do not provide high availability or fault tolerance. For information about creating a cluster with replication, see Creating a Redis Replication Group (p. 162). For information about adding a read replica to a cluster with replication, see Adding a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 200).

Topics
- Enabling Multi-AZ with Automatic Failover (Console) (p. 154)
- Enabling Multi-AZ with Automatic Failover (AWS CLI) (p. 154)
- Enabling Multi-AZ with Automatic Failover (ElastiCache API) (p. 155)

Enabling Multi-AZ with Automatic Failover (Console)

You can enable Multi-AZ with Automatic Failover using the ElastiCache console when you create a new Redis cluster or by modifying an existing Redis cluster with replication.

Multi-AZ with Automatic Failover is enabled by default and cannot be disabled on Redis (cluster mode enabled) clusters.

Enabling Multi-AZ with Automatic Failover When Creating a Cluster Using the ElastiCache Console

For more information on this process, see Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79). Be sure to have one or more replicas and enable Multi-AZ with Automatic Failover.

Enabling Multi-AZ with Automatic Failover on an Existing Cluster (Console)

For more information on this process, see Modifying a Cluster Using the AWS Management Console (p. 100).

Enabling Multi-AZ with Automatic Failover (AWS CLI)

The following code example uses the AWS CLI to enable Multi-AZ with Automatic Failover for the replication group redis12.

**Important**

The replication group redis12 must already exist and have at least one available read replica.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group \
    --replication-group-id redis12 \
    --automatic-failover-enabled \
    --apply-immediately
```

For Windows:

```bash
aws elasticache modify-replication-group ^
    --replication-group-id redis12 ^
    --automatic-failover-enabled ^
    --apply-immediately
```
The JSON output from this command should look something like this.

```json
{
   "ReplicationGroup": {
      "Status": "modifying",
      "Description": "One shard, two nodes",
      "NodeGroups": [
         {
            "Status": "modifying",
            "NodeGroupMembers": [
               {
                  "CurrentRole": "primary",
                  "PreferredAvailabilityZone": "us-west-2b",
                  "CacheNodeId": "0001",
                  "ReadEndpoint": {
                     "Port": 6379,
                     "Address": "redis12-001.v5r9dc.0001.usw2.cache.amazonaws.com"
                  },
                  "CacheClusterId": "redis12-001"
               },
               {
                  "CurrentRole": "replica",
                  "PreferredAvailabilityZone": "us-west-2a",
                  "CacheNodeId": "0001",
                  "ReadEndpoint": {
                     "Port": 6379,
                     "Address": "redis12-002.v5r9dc.0001.usw2.cache.amazonaws.com"
                  },
                  "CacheClusterId": "redis12-002"
               }
            ],
            "NodeGroupId": "0001",
            "PrimaryEndpoint": {
               "Port": 6379,
               "Address": "redis12.v5r9dc.ng.0001.usw2.cache.amazonaws.com"
            }
         }
      ],
      "ReplicationGroupId": "redis12",
      "SnapshotRetentionLimit": 1,
      "AutomaticFailover": "enabling",
      "SnapshotWindow": "07:00-08:00",
      "SnapshottingClusterId": "redis12-002",
      "MemberClusters": [
         "redis12-001",
         "redis12-002"
      ],
      "PendingModifiedValues": {}
   }
}
```

For more information, see these topics in the AWS CLI Command Reference:

- `create-cache-cluster`
- `create-replication-group`
- `modify-replication-group` in the AWS CLI Command Reference.

Enabling Multi-AZ with Automatic Failover (ElastiCache API)

The following code example uses the ElastiCache API to enable Multi-AZ with Automatic Failover for the replication group `redis12`. 

API Version 2015-02-02
Note
To use this example, the replication group redis12 must already exist and have at least one available read replica.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=ModifyReplicationGroup
&ApplyImmediately=true
&AutoFailover=true
&ReplicationGroupId=redis12
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20140401T192317Z
&X-Amz-Credential=<credential>
```

For more information, see these topics in the ElastiCache API Reference:

- CreateCacheCluster
- CreateReplicationGroup
- ModifyReplicationGroup
Testing Multi-AZ with Automatic Failover

After you enable Multi-AZ with Automatic Failover, you can test it using the ElastiCache console, the AWS CLI, and the ElastiCache API.

When testing, note the following:

- You can use this operation to test automatic failover on up to five shards (called node groups in the ElastiCache API and AWS CLI) in any rolling 24-hour period.
- If you call this operation on shards in different clusters (called replication groups in the API and CLI), you can make the calls concurrently.
- If you call this operation multiple times on different shards in the same Redis (cluster mode enabled) replication group, the first node replacement must complete before a subsequent call can be made.
- To determine whether the node replacement is complete you can check events using the Amazon ElastiCache console, the AWS CLI, or the ElastiCache API. Look for the following events related to automatic failover, listed here in order of likely occurrence:
  1. Replication group message: Test Failover API called for node group <node-group-id>
  2. Cache cluster message: Failover from master node <primary-node-id> to replica node <node-id> completed
  3. Replication group message: Failover from master node <primary-node-id> to replica node <node-id> completed
  4. Cache cluster message: Recovering cache nodes <node-id>
  5. Cache cluster message: Finished recovery for cache nodes <node-id>

For more information, see the following:

- DescribeEvents in the ElastiCache API Reference
- describe-events in the AWS CLI Command Reference.

Testing Automatic Failover

- Testing Automatic Failover Using the AWS Management Console (p. 157)
- Testing Automatic Failover Using the AWS CLI (p. 158)
- Testing Automatic Failover Using the ElastiCache API (p. 160)

Testing Automatic Failover Using the AWS Management Console

The following procedure walks you through testing automatic failover.

To test automatic failover

2. In the navigation pane, choose Redis.
3. From the list of Redis clusters, choose the box to the left of the cluster you want to test. This cluster must have at least one read replica node.
4. In the Details area, confirm that this cluster is Multi-AZ enabled. If the cluster is not Multi-AZ enabled, either choose a different cluster or modify this cluster to enable Multi-AZ. For more information, see Using the AWS Management Console (p. 100).
5. For Redis (cluster mode disabled), choose the cluster's name.
   For Redis (cluster mode enabled), do the following:
   a. Choose the cluster's name.
   b. On the Shards page, for the shard (called node group in the API and CLI) on which you want to test failover, choose the shard's name.
6. On the Nodes page, choose Failover Primary.
7. Choose Continue to fail over the primary, or Cancel to cancel the operation and not fail over the primary node.
   During the failover process, the console continues to show the node's status as available. To track the progress of your failover test, choose Events from the console navigation pane. On the Events tab, watch for events that indicate your failover has started (Test Failover API called) and completed (Recovery completed).

Testing Automatic Failover Using the AWS CLI

You can test automatic failover on any Multi-AZ with Automatic Failover enabled cluster using the AWS CLI operation test-failover.

Parameters

- --replication-group-id – Required. The replication group (on the console, cluster) that is to be tested.
- --node-group-id – Required. The name of the node group you want to test automatic failover on. You can test a maximum of five node groups in a rolling 24-hour period.

The following example uses the AWS CLI to test automatic failover on the node group redis00-0003 in the Redis (cluster mode enabled) cluster redis00.

Example Test Automatic Failover

For Linux, macOS, or Unix:

```bash
aws elasticache test-failover \
   --replication-group-id redis00 \
   --node-group-id redis00-0003
```
Output from the preceding command looks something like this.

```
{
    "ReplicationGroup": {
        "Status": "available",
        "Description": "1 shard, 3 nodes (1 + 2 replicas)",
        "NodeGroups": [
            {
                "Status": "available",
                "NodeGroupMembers": [
                    {
                        "CurrentRole": "primary",
                        "PreferredAvailabilityZone": "us-west-2c",
                        "CacheNodeId": "0001",
                        "ReadEndpoint": {
                            "Port": 6379,
                            "Address": "redis1x3-001.7ekv3t.0001.usw2.cache.amazonaws.com"
                        },
                        "CacheClusterId": "redis1x3-001"
                    },
                    {
                        "CurrentRole": "replica",
                        "PreferredAvailabilityZone": "us-west-2a",
                        "CacheNodeId": "0001",
                        "ReadEndpoint": {
                            "Port": 6379,
                            "Address": "redis1x3-002.7ekv3t.0001.usw2.cache.amazonaws.com"
                        },
                        "CacheClusterId": "redis1x3-002"
                    },
                    {
                        "CurrentRole": "replica",
                        "PreferredAvailabilityZone": "us-west-2b",
                        "CacheNodeId": "0001",
                        "ReadEndpoint": {
                            "Port": 6379,
                            "Address": "redis1x3-003.7ekv3t.0001.usw2.cache.amazonaws.com"
                        },
                        "CacheClusterId": "redis1x3-003"
                    }
                ],
                "NodeGroupId": "0001",
                "PrimaryEndpoint": {
                    "Port": 6379,
                    "Address": "redis1x3.7ekv3t.ng.0001.usw2.cache.amazonaws.com"
                }
            }
        ],
        "ClusterEnabled": false,
        "ReplicationGroupId": "redis1x3",
        "SnapshotRetentionLimit": 1,
        "AutomaticFailover": "enabled",
        "SnapshotWindow": "11:30-12:30",
        "SnapshottingClusterId": "redis1x3-002",
        "MemberClusters": [
            "redis1x3-001",
            "redis1x3-002",
            "redis1x3-003"
        ]
    }
}
```
To track the progress of your failover, use the AWS CLI `describe-events` operation.

For more information, see the following:

- `test-failover` in the *AWS CLI Command Reference*.
- `describe-events` in the *AWS CLI Command Reference*.

### Testing Automatic Failover Using the ElastiCache API

You can test automatic failover on any cluster enabled with Multi-AZ with Automatic Failover using the ElastiCache API operation `TestFailover`.

**Parameters**

- `ReplicationGroupId` – Required. The replication group (on the console, cluster) that is to be tested.
- `NodeGroupId` – Required. The name of the node group you want to test automatic failover on. You can test a maximum of five node groups in a rolling 24-hour period.

The following example tests automatic failover on the node group `redis00-0003` in the replication group (on the console, cluster) `redis00`.

**Example Testing Automatic Failover**

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=TestFailover
  &ReplicationGroupId=redis00
  &NodeGroupId=redis00
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20140401T192317Z
  &X-Amz-Credential=<credential>
```

To track the progress of your failover, use the ElastiCache API `DescribeEvents` operation.

For more information, see the following:

- `TestFailover` in the *ElastiCache API Reference*
- `DescribeEvents` in the *ElastiCache API Reference*
How Synchronization and Backup are Implemented

All supported versions of Redis support backup and synchronization between the primary and replica nodes. However, the way that backup and synchronization is implemented varies depending on the Redis version.

Redis Version 2.8.22 and Later

Redis replication, in versions 2.8.22 and later, choose between two methods. For more information, see Redis Versions Before 2.8.22 (p. 161) and Backup and Restore for ElastiCache for Redis (p. 213).

During the forkless process, if the write loads are heavy, writes to the cluster are delayed to ensure that you don't accumulate too many changes and thus prevent a successful snapshot.

Redis Versions Before 2.8.22

Redis backup and synchronization in versions before 2.8.22 is a three-step process.

1. Fork, and in the background process, serialize the cluster data to disk. This creates a point-in-time snapshot.
2. In the foreground, accumulate a change log in the client output buffer.
   
   Important
   
   If the change log exceeds the client output buffer size, the backup or synchronization fails. For more information, see Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464).

3. Finally, transmit the cache data and then the change log to the replica node.
Creating a Redis Replication Group

You have the following options for creating a cluster with replica nodes. One applies when you already have an available Redis (cluster mode disabled) cluster not associated with any cluster that has replicas to use as the primary node. The other applies when you need to create a primary node with the cluster and read replicas. Currently, a Redis (cluster mode enabled) cluster must be created from scratch.

Option 1: Creating a Replication Group Using an Available Redis (Cluster Mode Disabled) Cluster (p. 163)

Use this option to leverage an existing single-node Redis (cluster mode disabled) cluster. You specify this existing node as the primary node in the new cluster, and then individually add 1 to 5 read replicas to the cluster. If the existing cluster is active, read replicas synchronize with it as they are created. See Creating a Replication Group Using an Available Redis (Cluster Mode Disabled) Cluster (p. 163).

**Important**

You cannot create a Redis (cluster mode enabled) cluster using an existing cluster. To create a Redis (cluster mode enabled) cluster (API/CLI: replication group) using the ElastiCache console, see Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 83).

Option 2: Creating a Redis Replication Group from Scratch (p. 168)

Use this option if you don't already have an available Redis (cluster mode disabled) cluster to use as the cluster's primary node, or if you want to create a Redis (cluster mode enabled) cluster. See Creating a Redis Replication Group from Scratch (p. 168).
Creating a Replication Group Using an Available Redis (Cluster Mode Disabled) Cluster

An available cluster is an existing single-node Redis cluster. Currently, Redis (cluster mode enabled) does not support creating a cluster with replicas using an available single-node cluster. If you want to create a Redis (cluster mode enabled) cluster, see Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 175).

The following procedure can only be used if you have a Redis (cluster mode disabled) single-node cluster. This cluster's node becomes the primary node in the new cluster. If you do not have a Redis (cluster mode disabled) cluster that you can use as the new cluster's primary, see Creating a Redis Replication Group from Scratch (p. 168).

Creating a Replication Group Using an Available Redis Cluster (Console)

See the topic Using the AWS Management Console (p. 104).

Creating a Replication Group Using an Available Redis Cache Cluster (AWS CLI)

There are two steps to creating a replication group with read replicas when using an available Redis Cache Cluster for the primary when using the AWS CLI.

When using the AWS CLI you create a replication group specifying the available standalone node as the cluster's primary node, --primary-cluster-id and the number of nodes you want in the cluster using the CLI command, create-replication-group. Include the following parameters.

--replication-group-id

The name of the replication group you are creating. The value of this parameter is used as the basis for the names of the added nodes with a sequential 3-digit number added to the end of the --replication-group-id. For example, sample-repl-group-001.

Redis (cluster mode disabled) replication group 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.

--replication-group-description

Description of the replication group.

--num-cache-clusters

The number of nodes you want in this cluster. This value includes the primary node. This parameter has a maximum value of six.

--primary-cluster-id

The name of the available Redis (cluster mode disabled) cluster's node that you want to be the primary node in this replication group.

If you want to enable in-transit or at-rest encryption on this replication group, add either or both of the --transit-encryption-enabled or --at-rest-encryption-enabled parameters and meet the following conditions.

- Your cluster must be running Redis version 3.2.6 or 4.0.10.
- The replication group must be created in an Amazon VPC.
You must also include the parameter `--cache-subnet-group`.
You must also include the parameter `--auth-token` with the customer specified string value for your AUTH token (password) needed to perform operations on this replication group.

The following command creates the replication group `sample-repl-group` using the available Redis (cluster mode disabled) cluster `redis01` as the replication group's primary node. It creates 2 new nodes which are read replicas. The settings of `redis01` (that is, parameter group, security group, node type, engine version, etc.) will be applied to all nodes in the replication group.

For Linux, macOS, or Unix:

```bash
aws elasticache create-replication-group \
    --replication-group-id sample-repl-group \
    --replication-group-description "demo cluster with replicas" \
    --num-cache-clusters 3 \
    --primary-cluster-id redis01
```

For Windows:

```bash
aws elasticache create-replication-group ^
    --replication-group-id sample-repl-group ^
    --replication-group-description "demo cluster with replicas" ^
    --num-cache-clusters 3 ^
    --primary-cluster-id redis01
```

For additional information and parameters you might want to use, see the AWS CLI topic `create-replication-group`.

Next, add read replicas to the replication group

After the replication group is created, add one to five read replicas to it using the `create-cache-cluster` command, being sure to include the following parameters.

```bash
--cache-cluster-id
```

The name of the cluster you are adding to the replication group.

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.

```bash
--replication-group-id
```

The name of the replication group to which you are adding this cache cluster.

Repeat this command for each read replica you want to add to the replication group, changing only the value of the `--cache-cluster-id` parameter.

**Note**

Remember, a replication group cannot have more than five read replicas. Attempting to add a read replica to a replication group that already has five read replicas causes the operation to fail.

The following code adds the read replica `my-replica01` to the replication group `sample-repl-group`. The settings of the primary cluster parameter group, security group, node type, etc. will be applied to nodes as they are added to the replication group.
For Linux, macOS, or Unix:

```
aws elasticache create-cache-cluster \\n   --cache-cluster-id my-replica01 \\n   --replication-group-id sample-repl-group
```

For Windows:

```
aws elasticache create-cache-cluster ^
   --cache-cluster-id my-replica01 ^
   --replication-group-id sample-repl-group
```

Output from this command will look something like this.

```
{
   "ReplicationGroup": {
      "Status": "creating",
      "Description": "demo cluster with replicas",
      "ClusterEnabled": false,
      "ReplicationGroupId": "sample-repl-group",
      "SnapshotRetentionLimit": 1,
      "AutomaticFailover": "disabled",
      "SnapshotWindow": "00:00-01:00",
      "SnapshottingClusterId": "redis01",
      "MemberClusters": [
         "sample-repl-group-001",
         "sample-repl-group-002",
         "redis01"
      ],
      "CacheNodeType": "cache.m4.large",
      "PendingModifiedValues": {}  
   }
}
```

For additional information, see the AWS CLI topics:

- create-replication-group
- modify-replication-group

### Adding Replicas to a Standalone Redis (Cluster Mode Disabled) Cluster (ElastiCache API)

When using the ElastiCache API, you create a replication group specifying the available standalone node as the cluster's primary node, `PrimaryClusterId` and the number of nodes you want in the cluster using the CLI command, `CreateReplicationGroup`. Include the following parameters.

**ReplicationGroupId**

The name of the replication group you are creating. The value of this parameter is used as the basis for the names of the added nodes with a sequential 3-digit number added to the end of the `ReplicationGroupId`. For example, `sample-repl-group-001`.

Redis (cluster mode disabled) replication group 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.
ReplicationGroupDescription

Description of the cluster with replicas.

NumCacheClusters

The number of nodes you want in this cluster. This value includes the primary node. This parameter has a maximum value of six.

PrimaryClusterId

The name of the available Redis (cluster mode disabled) cluster that you want to be the primary node in this cluster.

If you want to enable in-transit or at-rest encryption on this replication group, add either or both of the TransitEncryptionEnabled=true or AtRestEncryptionEnabled=true parameters and meet the following conditions.

• Your cluster must be running Redis version 3.2.6, 4.0.10 or later.
• The replication group must be created in an Amazon VPC.
• You must also include the parameter CacheSubnetGroup.
• You must also include the parameter AuthToken with the customer specified string value for your AUTH token (password) needed to perform operations on this replication group.

The following command creates the cluster with replicas sample-repl-group using the available Redis (cluster mode disabled) cluster redis01 as the replication group's primary node. It creates 2 new nodes which are read replicas. The settings of redis01 (that is, parameter group, security group, node type, engine version, etc.) will be applied to all nodes in the replication group.

https://elasticache.us-west-2.amazonaws.com/
?Action=CreateReplicationGroup
&Engine=redis
&EngineVersion=3.2.4
&ReplicationGroupDescription=Demo%20cluster%20with%20replicas
&ReplicationGroupId=sample-repl-group
&PrimaryClusterId=redis01
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>

For additional information, see the ElastiCache APL topics:

- CreateReplicationGroup
- ModifyReplicationGroup

Next, add read replicas to the replication group

After the replication group is created, add one to five read replicas to it using the CreateCacheCluster operation, being sure to include the following parameters.

CacheClusterId

The name of the cluster you are adding to the replication group.

Cluster naming constraints are as follows:

• Must contain 1–40 alphanumeric characters or hyphens.
Creating a Replication Group

- Must begin with a letter.
- Can't contain two consecutive hyphens.
- Can't end with a hyphen.

**ReplicationGroupId**

The name of the replication group to which you are adding this cache cluster.

Repeat this operation for each read replica you want to add to the replication group, changing only the value of the `CacheClusterId` parameter.

The following code adds the read replica `myReplica01` to the replication group `myReplGroup` The settings of the primary cluster–parameter group, security group, node type, etc.–will be applied to nodes as they are added to the replication group.

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateCacheCluster
  &CacheClusterId=myReplica01
  &ReplicationGroupId=myReplGroup
  &SignatureMethod=HmacSHA256
  &SignatureVersion=4
  &Version=2015-02-02
  &X-Amz-Credential=[your-access-key-id]/20150202/us-west-2/elasticache/aws4_request
  &X-Amz-Date=20150202T170651Z
  &X-Amz-SignedHeaders=content-type;host;user-agent;x-amz-content-sha256;x-amz-date
  &X-Amz-Signature=[signature-value]
```

For additional information and parameters you might want to use, see the ElastiCache API topic CreateCacheCluster.
Creating a Redis Replication Group from Scratch

This topic covers how to create a Redis replication group without using an existing Redis cluster as the primary. You can create a Redis (cluster mode disabled) or Redis (cluster mode enabled) replication group from scratch using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Before you continue, decide whether you want to create a Redis (cluster mode disabled) or a Redis (cluster mode enabled) replication group. For guidance in deciding, see Replication: Redis (Cluster Mode Disabled) vs. Redis (Cluster Mode Enabled) (p. 145).

Topics
- Creating a Redis (cluster mode disabled) Replication Group from Scratch (p. 169)
- Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (p. 175)
Creating a Redis (cluster mode disabled) Replication Group from Scratch

You can create a Redis (cluster mode disabled) replication group from scratch using the ElastiCache console, the AWS CLI, or the ElastiCache API. A Redis (cluster mode disabled) replication group always has one node group, a primary cluster, and up to 5 read replicas. The configuration can range from 90 shards and 0 replicas to 15 shards and 5 replicas, which is the maximum number or replicas allowed. Redis (cluster mode disabled) replication groups do not support partitioning your data.

Note
The node/shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see AWS Service Limits and include the instance type in the request.

Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch

- Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79) Specify at least one replica node in step 6.i.
- Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch (AWS CLI) (p. 169)
- Creating a Redis (cluster mode disabled) Replication Group from Scratch (ElastiCache API) (p. 172)

Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch (AWS CLI)

The following procedure creates a Redis (cluster mode disabled) replication group using the AWS CLI.

When you create a Redis (cluster mode disabled) replication group from scratch, you create the replication group and all its nodes with a single call to the AWS CLI `create-replication-group` command. Include the following parameters.

```bash
--replication-group-id
```

The name of the replication group you are creating.

Redis (cluster mode disabled) replication group 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.

```bash
--replication-group-description
```

Description of the replication group.

```bash
--num-cache-clusters
```

The number of nodes you want created with this replication group, primary and read replicas combined.

If you enable Multi-AZ (`--automatic-failover-enabled`), the value of `--num-cache-clusters` must be at least 2.

```bash
--cache-node-type
```

The node type for each node in the replication group.

The following node types are supported by ElastiCache. 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

**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:

### Memory optimized:

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

### Compute optimized:

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

### Memory optimized:

- **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

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

### Additional node type info

- All current generation instance types are created in Amazon VPC by default.
- Redis append-only files (AOF) are not supported for T1 or T2 instances.
- Redis Multi-AZ with automatic failover is not supported on T1 instances.
- Redis configuration variables appendonly and appendfsync are not supported on Redis version 2.8.22 and later.

```
--cache-parameter-group
```

Specify a parameter group that corresponds to your engine version. If you are running Redis 3.2.4 or later, specify the default.redis3.2 parameter group or a parameter group derived from default.redis3.2 to create a Redis (cluster mode disabled) replication group. For more information, see Redis Specific Parameters (p. 315).

```
--engine
redis
--engine-version
```

To have the richest set of features, choose the latest engine version.
The names of the nodes will be derived from the replication group name by postpending -00# to the replication group name. For example, using the replication group name myReplGroup, the name for the primary will be myReplGroup-001 and the read replicas myReplGroup-002 through myReplGroup-006.

If you want to enable in-transit or at-rest encryption on this replication group, add either or both of the --transit-encryption-enabled or --at-rest-encryption-enabled parameters and meet the following conditions.

- Your replication group must be running Redis version 3.2.6 or 4.0.10.
- The replication group must be created in an Amazon VPC.
- You must also include the parameter --cache-subnet-group.
- You must also include the parameter --auth-token with the customer specified string value for your AUTH token (password) needed to perform operations on this replication group.

The following operation creates a Redis (cluster mode disabled) replication group sample-repl-group with three nodes, a primary and two replicas.

For Linux, macOS, or Unix:

```bash
aws elasticache create-replication-group
  --replication-group-id sample-repl-group
  --replication-group-description "Demo cluster with replicas"
  --num-cache-clusters 3
  --cache-node-type cache.m4.large
  --cache-parameter-group default.redis3.2
  --engine redis
  --engine-version 3.2.4
```

For Windows:

```bash
aws elasticache create-replication-group
  --replication-group-id sample-repl-group
  --replication-group-description "Demo cluster with replicas"
  --num-cache-clusters 3
  --cache-node-type cache.m4.large
  --cache-parameter-group default.redis3.2
  --engine redis
  --engine-version 3.2.4
```

Output from the this command is something like this.

```json
{
  "ReplicationGroup": {
    "Status": "creating",
    "Description": "Demo cluster with replicas",
    "ClusterEnabled": false,
    "ReplicationGroupId": "sample-repl-group",
    "SnapshotRetentionLimit": 0,
    "AutomaticFailover": "disabled",
    "SnapshotWindow": "01:30-02:30",
    "MemberClusters": [
      "sample-repl-group-001",
      "sample-repl-group-002",
      "sample-repl-group-003"
    ],
    "CacheNodeType": "cache.m4.large",
    "PendingModifiedValues": {}
  }
}
```
For additional information and parameters you might want to use, see the AWS CLI topic `create-replication-group`.

Creating a Redis (cluster mode disabled) Replication Group from Scratch (ElastiCache API)

The following procedure creates a Redis (cluster mode disabled) replication group using the ElastiCache API.

When you create a Redis (cluster mode disabled) replication group from scratch, you create the replication group and all its nodes with a single call to the ElastiCache API `CreateReplicationGroup` operation. Include the following parameters.

**ReplicationGroupId**

The name of the replication group you are creating.

Redis (cluster mode enabled) replication group 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.

**ReplicationGroupDescription**

Your description of the replication group.

**NumCacheClusters**

The total number of nodes you want created with this replication group, primary and read replicas combined.

If you enable Multi-AZ (`AutomaticFailoverEnabled=true`), the value of `NumCacheClusters` must be at least 2.

**CacheNodeType**

The node type for each node in the replication group.

The following node types are supported by ElastiCache. 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
    - **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

Additional node type info

- All current generation instance types are created in Amazon VPC by default.
- Redis append-only files (AOF) are not supported for T1 or T2 instances.
- Redis Multi-AZ with automatic failover is not supported on T1 instances.
- Redis configuration variables appendonly and appendfsync are not supported on Redis version 2.8.22 and later.

CacheParameterGroup

Specify a parameter group that corresponds to your engine version. If you are running Redis 3.2.4 or later, specify the default.redis3.2 parameter group or a parameter group derived from default.redis3.2 to create a Redis (cluster mode disabled) replication group. For more information, see Redis Specific Parameters (p. 315).

Engine

redis

EngineVersion

3.2.4

The names of the nodes will be derived from the replication group name by postpending -00# to the replication group name. For example, using the replication group name myReplGroup, the name for the primary will be myReplGroup-001 and the read replicas myReplGroup-002 through myReplGroup-006.

If you want to enable in-transit or at-rest encryption on this replication group, add either or both of the TransitEncryptionEnabled=true or AtRestEncryptionEnabled=true parameters and meet the following conditions.

- Your replication group must be running Redis version 3.2.6 or 4.0.10.
- The replication group must be created in an Amazon VPC.
- You must also include the parameter CacheSubnetGroup.
- You must also include the parameter AuthToken with the customer specified string value for your AUTH token (password) needed to perform operations on this replication group.
The following operation creates the Redis (cluster mode disabled) replication group myReplGroup with three nodes, a primary and two replicas.

https://elasticache.us-west-2.amazonaws.com/
?Action=CreateReplicationGroup
&CacheNodeType=cache.m4.large
&CacheParameterGroup=default.redis3.2
&Engine=redis
&EngineVersion=3.2.4
&NumCacheClusters=3
&ReplicationGroupDescription=test%20group
&ReplicationGroupId=myReplGroup
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>

For additional information and parameters you might want to use, see the ElastiCache API topic CreateReplicationGroup.
Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch

You can create a Redis (cluster mode enabled) cluster (API/CLI: replication group) using the ElastiCache console, the AWS CLI, or the ElastiCache API. A Redis (cluster mode enabled) replication group has from 1 to 90 shards (API/CLI: node groups), a primary node in each shard, and up to 5 read replicas in each shard. You can create a cluster with higher number of shards and lower number of replicas totaling up to 90 nodes per cluster. This cluster configuration can range from 90 shards and 0 replicas to 15 shards and 5 replicas, which is the maximum number of replicas allowed.

Note
The node or shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see AWS Service Limits and select limit type "Nodes per cluster per instance type".

Creating a Cluster in Redis (Cluster Mode Enabled)

- Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 175)
- Creating a Redis (Cluster Mode Enabled) Replication Group from Scratch (AWS CLI) (p. 175)
- Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (ElastiCache API) (p. 179)

Creating a Redis (Cluster Mode Enabled) Cluster (Console)

To create a Redis (cluster mode enabled) cluster, see Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 83). Be sure to enable cluster mode, Cluster Mode enabled (Scale Out), and specify at least two shards and one replica node in each.

Creating a Redis (Cluster Mode Enabled) Replication Group from Scratch (AWS CLI)

The following procedure creates a Redis (cluster mode enabled) replication group using the AWS CLI.

When you create a Redis (cluster mode enabled) replication group from scratch, you create the replication group and all its nodes with a single call to the AWS CLI create-replication-group command. Include the following parameters.

--replication-group-id
The name of the replication group you are creating.

Redis (cluster mode enabled) replication group 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.

--replication-group-description
Description of the replication group.

--cache-node-type
The node type for each node in the replication group.

The following node types are supported by ElastiCache. 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

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

Additional node type info
  • All current generation instance types are created in Amazon VPC by default.
  • Redis append-only files (AOF) are not supported for T1 or T2 instances.
  • Redis Multi-AZ with automatic failover is not supported on T1 instances.
  • Redis configuration variables appendonly and appendfsync are not supported on Redis version 2.8.22 and later.

--cache-parameter-group

Specify the default.redis3.2.cluster.on parameter group or a parameter group derived from default.redis3.2.cluster.on to create a Redis (cluster mode enabled) replication group. For more information, see Redis 3.2.4 Parameter Changes (p. 323).

--engine

redis

--engine-version

3.2.4

--num-node-groups

The number of node groups in this replication group. Valid values are 1 to 90.

Note

The node/shard limit can be increased to a maximum of 250 per cluster. To request a limit increase, see AWS Service Limits and select limit type "Nodes per cluster per instance type".
The number of replica nodes in each node group. Valid values are 0 to 5.

If you want to enable in-transit or at-rest encryption on this replication group, add either or both of the
--transit-encryption-enabled or --at-rest-encryption-enabled parameters and meet the following conditions.

- Your replication group must be running Redis version 3.2.6 or 4.0.10.
- The replication group must be created in an Amazon VPC.
- You must also include the parameter --cache-subnet-group.
- You must also include the parameter --auth-token with the customer specified string value for your
  AUTH token (password) needed to perform operations on this replication group.

The following operation creates the Redis (cluster mode enabled) replication group sample-repl-
group with three node groups/shards (--num-node-groups), each with three nodes, a primary and two
read replicas (--replicas-per-node-group).

For Linux, macOS, or Unix:

```bash
aws elasticache create-replication-group \
  --replication-group-id sample-repl-group \
  --replication-group-description "Demo cluster with replicas" \
  --num-node-groups 3 \
  --replicas-per-node-group 2 \
  --cache-node-type cache.m4.large \
  --cache-parameter-group default.redis3.2.cluster.on \
  --engine redis \
  --engine-version 3.2.4 \
  --security-group-ids SECURITY_GROUP_ID \
  --cache-subnet-group-name SUBNET_GROUP_NAME>
```

For Windows:

```bash
aws elasticache create-replication-group ^
  --replication-group-id sample-repl-group ^
  --replication-group-description "Demo cluster with replicas" ^
  --num-node-groups 3 ^
  --replicas-per-node-group 2 ^
  --cache-node-type cache.m4.large ^
  --cache-parameter-group default.redis3.2.cluster.on ^
  --engine redis ^
  --engine-version 3.2.4 ^
  --security-group-ids SECURITY_GROUP_ID ^
  --cache-subnet-group-name SUBNET_GROUP_NAME>
```

The preceding command generates the following output.

```json
{
  "ReplicationGroup": {
    "Status": "creating",
    "Description": "Demo cluster with replicas",
    "ReplicationGroupId": "sample-repl-group",
    "SnapshotRetentionLimit": 0,
    "AutomaticFailover": "enabled",
    "SnapshotWindow": "05:30-06:30"
  }
}
```
When you create a Redis (cluster mode enabled) replication group from scratch, you are able to configure each shard in the cluster using the `--node-group-configuration` parameter as shown in the following example which configures two node groups (Console: shards). The first shard has two nodes, a primary and one read replica. The second shard has three nodes, a primary and two read replicas.

```
When you create a Redis (cluster mode enabled) replication group from scratch, you are able to configure each shard in the cluster using the `--node-group-configuration` parameter as shown in the following example which configures two node groups (Console: shards). The first shard has two nodes, a primary and one read replica. The second shard has three nodes, a primary and two read replicas.

--node-group-configuration

The configuration for each node group. The `--node-group-configuration` parameter consists of the following fields.

- **PrimaryAvailabilityZone** – The Availability Zone where the primary node of this node group is located. If this parameter is omitted, ElastiCache chooses the Availability Zone for the primary node.

  **Example:** us-west-2a.

- **ReplicaAvailabilityZones** – A comma separated list of Availability Zones where the read replicas are located. The number of Availability Zones in this list must match the value of `ReplicaCount`. If this parameter is omitted, ElastiCache chooses the Availability Zones for the replica nodes.

  **Example:** "us-west-2a,us-west-2b,us-west-2c"

- **ReplicaCount** – The number of replica nodes in this node group.

- **Slots** – A string that specifies the keyspace for the node group. The string is in the format `startKey-endKey`. If this parameter is omitted, ElastiCache allocates keys equally among the node groups.

  **Example:** "0-4999"

The following operation creates the Redis (cluster mode enabled) replication group `new-group` with two node groups/shards (`--num-node-groups`). Unlike the preceding example, each node group is configured differently from the other node group (`--node-group-configuration`).

For Linux, macOS, or Unix:

```
aws elasticache create-replication-group \
  --replication-group-id new-group \
  --replication-group-description "Sharded replication group" \
  --engine redis \
  --engine-version 3.2.4 \
  --cache-parameter-group default.redis3.2.cluster.on \
  --snapshot-retention-limit 8 \
  --cache-node-type cache.m4.medium \
  --num-node-groups 2 \
```
```
Creating a Replication Group

```
--node-group-configuration \
   "ReplicaCount=1,Slots=0-8999,PrimaryAvailabilityZone='us-east-1c',ReplicaAvailabilityZones='us-east-1b' \ 
   "ReplicaCount=2,Slots=9000-16383,PrimaryAvailabilityZone='us-east-1a',ReplicaAvailabilityZones='us-east-1a','us-east-1c'"
```

For Windows:

```
aws elasticache create-replication-group ^
   --replication-group-id new-group ^
   --replication-group-description "Sharded replication group" ^
   --engine redis ^
   --engine-version 3.2.4 ^
   --cache-parameter-group default.redis3.2.cluster.on ^
   --snapshot-retention-limit 8 ^
   --cache-node-type cache.m4.medium ^
   --num-node-groups 2 ^
   --node-group-configuration \ 
   "ReplicaCount=1,Slots=0-8999,PrimaryAvailabilityZone='us-east-1c',ReplicaAvailabilityZones='us-east-1b' \ 
   "ReplicaCount=2,Slots=9000-16383,PrimaryAvailabilityZone='us-east-1a',ReplicaAvailabilityZones='us-east-1a','us-east-1c'"
```

The preceding operation generates the following output.

```
{
   "ReplicationGroup": {
      "Status": "creating",
      "Description": "Sharded replication group",
      "ReplicationGroupId": "rc-rg",
      "SnapshotRetentionLimit": 8,
      "AutomaticFailover": "enabled",
      "SnapshotWindow": "10:00-11:00",
      "MemberClusters": [
         "rc-rg-0001-001",
         "rc-rg-0001-002",
         "rc-rg-0002-001",
         "rc-rg-0002-002",
         "rc-rg-0002-003"
      ],
      "PendingModifiedValues": {}
   }
}
```

For additional information and parameters you might want to use, see the AWS CLI topic `create-replication-group`.

Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (ElastiCache API)

The following procedure creates a Redis (cluster mode enabled) replication group using the ElastiCache API.

When you create a Redis (cluster mode enabled) replication group from scratch, you create the replication group and all its nodes with a single call to the ElastiCache API `CreateReplicationGroup` operation. Include the following parameters.

**ReplicationGroupId**

The name of the replication group you are creating.

Redis (cluster mode enabled) replication group 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.

ReplicationGroupDescription

Description of the replication group.

NumNodeGroups

The number of node groups you want created with this replication group. Valid values are 1 to 90.

ReplicasPerNodeGroup

The number of replica nodes in each node group. Valid values are 1 to 5.

NodeGroupConfiguration

The configuration for each node group. The NodeGroupConfiguration parameter consists of the following fields.
• PrimaryAvailabilityZone – The Availability Zone where the primary node of this node group is located. If this parameter is omitted, ElastiCache chooses the Availability Zone for the primary node.
  
  Example: us-west-2a.

• ReplicaAvailabilityZones – A list of Availability Zones where the read replicas are located. The number of Availability Zones in this list must match the value of ReplicaCount. If this parameter is omitted, ElastiCache chooses the Availability Zones for the replica nodes.

• ReplicaCount – The number of replica nodes in this node group.

• Slots – A string that specifies the keyspace for the node group. The string is in the format startKey-endKey. If this parameter is omitted, ElastiCache allocates keys equally among the node groups.
  
  Example: "0-4999"

CacheNodeType

The node type for each node in the replication group.

The following node types are supported by ElastiCache. 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

    **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

Additional node type info

- All current generation instance types are created in Amazon VPC by default.
- Redis append-only files (AOF) are not supported for T1 or T2 instances.
- Redis Multi-AZ with automatic failover is not supported on T1 instances.
- Redis configuration variables appendonly and appendfsync are not supported on Redis version 2.8.22 and later.

CacheParameterGroup

Specify the default.redis3.2.cluster.on parameter group or a parameter group derived from default.redis3.2.cluster.on to create a Redis (cluster mode enabled) replication group. For more information, see Redis 3.2.4 Parameter Changes (p. 323).

Engine

redis

EngineVersion

3.2.4

If you want to enable in-transit or at-rest encryption on this replication group, add either or both of the TrasitEncryptionEnabled=true or AtRestEncryptionEnabled=true parameters and meet the following conditions.

- Your replication group must be running Redis version 3.2.6 or 4.0.10.
- The replication group must be created in an Amazon VPC.
- You must also include the parameter CacheSubnetGroup.
- You must also include the parameter AuthToken with the customer specified string value for your AUTH token (password) needed to perform operations on this replication group.

Line breaks are added for ease of reading.
Viewing a Replication Group's Details

There are times you may want to view the details of a replication group. You can use the ElastiCache console, the AWS CLI for ElastiCache, or the ElastiCache API. The console process is different for Redis (cluster mode disabled) and Redis (cluster mode enabled).

Viewing a Replication Group's Details

• Viewing Details for a Redis (Cluster Mode Disabled) with Replicas (p. 182)
  • Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (Console) (p. 182)
  • Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (AWS CLI) (p. 183)
  • Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (ElastiCache API) (p. 183)
• Viewing a Replication Group's Details: Redis (Cluster Mode Enabled) (p. 183)
  • Viewing Details for a Redis (Cluster Mode Enabled) Cluster (Console) (p. 183)
  • Viewing Details for a Redis (Cluster Mode Enabled) Cluster (AWS CLI) (p. 183)
  • Viewing Details for a Redis (Cluster Mode Enabled) Cluster (ElastiCache API) (p. 183)
• Viewing a Replication Group's Details (AWS CLI) (p. 183)
• Viewing a Replication Group's Details (ElastiCache API) (p. 185)

Viewing Details for a Redis (Cluster Mode Disabled) with Replicas

You can view the details of a Redis (cluster mode disabled) cluster with replicas (API/CLI: replication group) using the ElastiCache console, the AWS CLI for ElastiCache, or the ElastiCache API.

Viewing a Redis (Cluster Mode Disabled) Cluster's Details

• Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (Console) (p. 182)
• Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (AWS CLI) (p. 183)
• Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (ElastiCache API) (p. 183)

Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (Console)

To view the details of a Redis (cluster mode disabled) cluster with replicas using the ElastiCache console, see the topic Viewing a Redis (Cluster Mode Disabled) Cluster's Details (Console) (p. 90).
Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (AWS CLI)

For an AWS CLI example that displays a Redis (cluster mode disabled) replication group's details, see Viewing a Replication Group's Details (AWS CLI) (p. 183).

Viewing Details for a Redis (Cluster Mode Disabled) Replication Group (ElastiCache API)

For an ElastiCache API example that displays a Redis (cluster mode disabled) replication group's details, see Viewing a Replication Group's Details (ElastiCache API) (p. 185).

Viewing a Replication Group's Details: Redis (Cluster Mode Enabled)

Viewing Details for a Redis (Cluster Mode Enabled) Cluster (Console)

To view the details of a Redis (cluster mode enabled) cluster using the ElastiCache console, see Viewing Details for a Redis (Cluster Mode Enabled) Cluster (Console) (p. 91).

Viewing Details for a Redis (Cluster Mode Enabled) Cluster (AWS CLI)

For an ElastiCache CLI example that displays a Redis (cluster mode enabled) replication group's details, see Viewing a Replication Group's Details (AWS CLI) (p. 183).

Viewing Details for a Redis (Cluster Mode Enabled) Cluster (ElastiCache API)

For an ElastiCache API example that displays a Redis (cluster mode enabled) replication group's details, see Viewing a Replication Group's Details (ElastiCache API) (p. 185).

Viewing a Replication Group's Details (AWS CLI)

You can view the details for a replication using the AWS CLI `describe-global-replication-groups` command. Use the following optional parameters to refine the listing. Omitting the parameters returns the details for up to 100 replication groups.

**Optional Parameters**

- `--global-replication-group-id` – Use this parameter to list the details of a specific replication group. If the specified replication group has more than one node group, results are returned grouped by node group.
- `--max-items` – Use this parameter to limit the number of replication groups listed. The value of `--max-items` cannot be less than 20 or greater than 100.

**Example**

The following code lists the details for up to 100 replication groups.

```
aws elasticache describe-replication-groups
```

The following code lists the details for `sample-repl-group`.

```
aws elasticache describe-replication-groups --replication-group-id sample-repl-group
```

The following code lists the details for `sample-repl-group`.

```
aws elasticache describe-replication-groups --replication-group-id sample-repl-group
```
The following code list the details for up to 25 replication groups.

```
aws elasticache describe-replication-groups --replication-group-id sample-repl-group
```

Output from this operation should look something like this (JSON format).

```
{
  "ReplicationGroups": [
    {
      "Status": "available",
      "Description": "test",
      "NodeGroups": [
        {
          "Status": "available",
          "NodeGroupMembers": [
            {
              "CurrentRole": "primary",
              "PreferredAvailabilityZone": "us-west-2a",
              "CacheNodeId": "0001",
              "ReadEndpoint": {
                "Port": 6379,
                "Address": "rg-name-001.1abc4d.0001.usw2.cache.amazonaws.com"
              },
              "CacheClusterId": "rg-name-001"
            },
            {
              "CurrentRole": "replica",
              "PreferredAvailabilityZone": "us-west-2b",
              "CacheNodeId": "0001",
              "ReadEndpoint": {
                "Port": 6379,
                "Address": "rg-name-002.1abc4d.0001.usw2.cache.amazonaws.com"
              },
              "CacheClusterId": "rg-name-002"
            },
            {
              "CurrentRole": "replica",
              "PreferredAvailabilityZone": "us-west-2c",
              "CacheNodeId": "0001",
              "ReadEndpoint": {
                "Port": 6379,
                "Address": "rg-name-003.1abc4d.0001.usw2.cache.amazonaws.com"
              },
              "CacheClusterId": "rg-name-003"
            }
          ],
          "NodeGroupId": "0001",
          "PrimaryEndpoint": {
            "Port": 6379,
            "Address": "rg-name.1abc4d.ng.0001.usw2.cache.amazonaws.com"
          }
        }
      ],
      "ReplicationGroupId": "rg-name",
      "AutomaticFailover": "enabled",
      "SnapshottingClusterId": "rg-name-002",
      "MemberClusters": [
        "rg-name-001",
        "rg-name-002",
        "rg-name-003"
      ]
    }
  ],
  "API Version": "2015-02-02"
}
```

184
"PendingModifiedValues": {}
},
{
... some output omitted for brevity
}
]

For more information, see the AWS CLI for ElastiCache topic `describe-replication-groups`.

**Viewing a Replication Group's Details (ElastiCache API)**

You can view the details for a replication using the AWS CLI `DescribeReplicationGroups` operation. Use the following optional parameters to refine the listing. Omitting the parameters returns the details for up to 100 replication groups.

**Optional Parameters**

- **ReplicationGroupId** – Use this parameter to list the details of a specific replication group. If the specified replication group has more than one node group, results are returned grouped by node group.
- **MaxRecords** – Use this parameter to limit the number of replication groups listed. The value of `MaxRecords` cannot be less than 20 or greater than 100. The default is 100.

**Example**

The following code list the details for up to 100 replication groups.

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

The following code lists the details for `myReplGroup`.

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

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

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeReplicationGroups
&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 `DescribeReplicationGroups`. 
Finding Replication Group Endpoints

An application can connect to any node in a replication group, provided that it has the DNS endpoint and port number for that node. Depending upon whether you are running a Redis (cluster mode disabled) or a Redis (cluster mode enabled) replication group, you will be interested in different endpoints.

Redis (Cluster Mode Disabled)

Redis (cluster mode disabled) clusters with replicas have three types of endpoints; the primary endpoint, the reader endpoint and the node endpoints. The primary endpoint is a DNS name that always resolves to the primary node in the cluster. The primary endpoint is immune to changes to your cluster, such as promoting a read replica to the primary role. For write activity, we recommend that your applications connect to the primary endpoint instead of connecting directly to the primary.

A reader endpoint will evenly split incoming connections to the endpoint between all read replicas in an ElastiCache for Redis cluster. Reader endpoints keep up with cluster changes in real-time as replicas are added or removed. You can place your ElastiCache for Redis cluster's multiple read replicas in different AWS Availability Zones (AZ) to ensure high availability of reader endpoints.

For read activity, applications can also connect to any node in the cluster. Unlike the primary endpoint, node endpoints resolve to specific endpoints. If you make a change in your cluster, such as adding or deleting a replica, you must update the node endpoints in your application.

Redis (Cluster Mode Enabled)

Redis (cluster mode enabled) clusters with replicas, because they have multiple shards (API/CLI: node groups), which mean they also have multiple primary nodes, have a different endpoint structure than Redis (cluster mode disabled) clusters. Redis (cluster mode enabled) has a configuration endpoint which "knows" all the primary and node endpoints in the cluster. Your application connects to the configuration endpoint. Whenever your application writes to or reads from the cluster's configuration endpoint, Redis, behind the scenes, determines which shard the key belongs to and which endpoint in that shard to use. It is all quite transparent to your application.

You can find the endpoints for a cluster using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Finding Replication Group Endpoints

To find the endpoints for your replication group, see one of the following topics:

- Finding a Redis (Cluster Mode Disabled) Cluster's Endpoints (Console) (p. 206)
- Finding Endpoints for a Redis (Cluster Mode Enabled) Cluster (Console) (p. 208)
- Finding the Endpoints for Replication Groups (AWS CLI) (p. 211)
- Finding Endpoints for Replication Groups (ElastiCache API) (p. 213)
Modifying a Replication Group

Important Constraints

- Currently, ElastiCache supports limited modifications of a Redis (cluster mode enabled) replication group, for example changing the engine version, using the API operation `ModifyReplicationGroup` (CLI: `modify-replication-group`). You can modify the number of shards (node groups) in a Redis (cluster mode enabled) cluster with the API operation `ModifyReplicationGroupShardConfiguration` (CLI: `modify-replication-group-shard-configuration`). For more information, see Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279).

Other modifications to a Redis (cluster mode enabled) cluster require that you create a cluster with the new cluster incorporating the changes.

- You can upgrade Redis (cluster mode disabled) and Redis (cluster mode enabled) clusters and replication groups to newer engine versions. However, you can't downgrade to earlier engine versions except by deleting the existing cluster or replication group and creating it again. For more information, see Upgrading Engine Versions (p. 53).

You can modify a Redis (cluster mode disabled) cluster's settings using the ElastiCache console, the AWS CLI, or the ElastiCache API. Currently, ElastiCache supports a limited number of modifications on a Redis (cluster mode enabled) replication group. Other modifications require you create a backup of the current replication group then using that backup to seed a new Redis (cluster mode enabled) replication group.

Topics

- Using the AWS Management Console (p. 188)
- Using the AWS CLI (p. 188)
- Using the ElastiCache API (p. 189)

Using the AWS Management Console

To modify a Redis (cluster mode disabled) cluster, see Modifying an ElastiCache Cluster (p. 100).

Using the AWS CLI

The following AWS CLI command enables Multi-AZ on an existing Redis replication group. You can use the same command to make other modifications to a replication group.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group \
  --replication-group-id myReplGroup \
  --automatic-failover-enabled
```

For Windows:

```bash
aws elasticache modify-replication-group ^
  --replication-group-id myReplGroup ^
  --automatic-failover-enabled
```

For more information on the AWS CLI `modify-replication-group` command, see `modify-replication-group`. 
Using the ElastiCache API

The following ElastiCache API operation enables Multi-AZ on an existing Redis replication group. You can use the same operation to make other modifications to a replication group.

https://elasticache.us-west-2.amazonaws.com/
?Action=ModifyReplicationGroup
&AutomaticFailoverEnabled=true
&ReplicationGroupId=myReplGroup
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&Version=2014-12-01
&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 on the ElastiCache API ModifyReplicationGroup operation, see ModifyReplicationGroup.
Deleting a Replication Group

If you no longer need one of your clusters with replicas (called replication groups in the API/CLI), you can delete it. When you delete a replication group, ElastiCache deletes all of the nodes in that group.

Once you have begun this operation, it cannot be interrupted or canceled.

**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.

Deleting a Replication Group (Console)

To delete a cluster that has replicas, see Deleting a Cluster (p. 117).

Deleting a Replication Group (AWS CLI)

Use the command `delete-replication-group` to delete a replication group.

```bash
aws elasticache delete-replication-group --replication-group-id my-repgroup
```

A prompt asks you to confirm your decision. Enter `y` (yes) to start the operation immediately. After the process starts, it is irreversible.

```
After you begin deleting this replication group, all of its nodes will be deleted as well.
Are you sure you want to delete this replication group? [Ny]y
REPLICATIONGROUP  my-repgroup  My replication group  deleting
```

Deleting a Replication Group (ElastiCache API)

Call `DeleteReplicationGroup` with the `ReplicationGroup` parameter.

**Example**

```bash
https://elasticache.us-west-2.amazonaws.com/
?Action=DeleteReplicationGroup
&ReplicationGroupId=my-repgroup
&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>
```

**Note**
If you set the `RetainPrimaryCluster` parameter to `true`, all of the read replicas will be deleted, but the primary cluster will be retained.
Changing the Number of Replicas

You can dynamically increase or decrease the number of read replicas in your Redis replication group using the AWS Management Console, the AWS CLI, or the ElastiCache API. If your replication group is a Redis (cluster mode enabled) replication group, you can choose which shards (node groups) to increase or decrease the number of replicas in.

To dynamically change the number of replicas in your Redis replication group, choose the operation from the following table that fits your situation.

<table>
<thead>
<tr>
<th>To Do This</th>
<th>For Redis (cluster mode enabled)</th>
<th>For Redis (cluster mode disabled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add replicas</td>
<td>Increasing the Number of Replicas in a Shard (p. 192)</td>
<td>Increasing the Number of Replicas in a Shard (p. 192)</td>
</tr>
<tr>
<td></td>
<td>Adding a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 200)</td>
<td></td>
</tr>
<tr>
<td>Delete replicas</td>
<td>Decreasing the Number of Replicas in a Shard (p. 196)</td>
<td>Decreasing the Number of Replicas in a Shard (p. 196)</td>
</tr>
<tr>
<td></td>
<td>Deleting a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 202)</td>
<td></td>
</tr>
</tbody>
</table>
Increasing the Number of Replicas in a Shard

You can increase the number of replicas in a Redis (cluster mode enabled) shard or Redis (cluster mode disabled) replication group up to a maximum of five. You can do so using the AWS Management Console, the AWS CLI, or the ElastiCache API.

Topics
- Using the AWS Management Console (p. 192)
- Using the AWS CLI (p. 192)
- Using the ElastiCache API (p. 194)

Using the AWS Management Console

The following procedure uses the console to increase the number of replicas in a Redis (cluster mode enabled) replication group.

To increase the number of replicas in Redis shards
2. In the navigation pane, choose Redis, and then choose the name of the replication group that you want to add replicas to.
3. Choose the box for each shard that you want to add replicas to.
4. Choose Add replicas.
5. Complete the Add Replicas to Shards page:
   - For New number of replicas/shard, enter the number of replicas that you want all of your selected shards to have. This value must be greater than or equal to Current Number of Replicas per shard and less than or equal to five. We recommend at least two replicas as a working minimum.
   - For Availability Zones, choose either No preference to have ElastiCache chose an Availability Zone for each new replica, or Specify Availability Zones to choose an Availability Zone for each new replica.
     If you choose Specify Availability Zones, for each new replica specify an Availability Zone using the list.
6. Choose Add to add the replicas or Cancel to cancel the operation.

Using the AWS CLI

To increase the number of replicas in a Redis shard, use the increase-replica-count command with the following parameters:

- --replication-group-id – Required. Identifies which replication group you want to increase the number of replicas in.
- --apply-immediately or --no-apply-immediately – Required. Specifies whether to increase the replica count immediately (--apply-immediately) or at the next maintenance window (--no-apply-immediately). Currently, --no-apply-immediately is not supported.
- --new-replica-count – Optional. Specifies the number of replica nodes you want when finished, up to a maximum of five. Use this parameter for Redis (cluster mode disabled) replication groups where there is only one node group or Redis (cluster mode enabled) group, or where you want all node groups to have the same number of replicas. If this value is not larger than the current number of replicas in the node group, the call fails with an exception.
• **--replica-configuration** – Optional. Allows you to set the number of replicas and Availability Zones for each node group independently. Use this parameter for Redis (cluster mode enabled) groups where you want to configure each node group independently.

--replica-configuration has three optional members:

• **NodeGroupId** – The four-digit ID for the node group that you are configuring. For Redis (cluster mode disabled) replication groups, the shard ID is always 0001. To find a Redis (cluster mode enabled) node group's (shard's) ID, see Finding a Shard's ID (p. 128).

• **NewReplicaCount** – The number of replicas that you want in this node group at the end of this operation. The value must be more than the current number of replicas, up to a maximum of five. If this value is not larger than the current number of replicas in the node group, the call fails with an exception.

• **PreferredAvailabilityZones** – A list of PreferredAvailabilityZone strings that specify which Availability Zones the replication group's nodes are to be in. The number of PreferredAvailabilityZone values must equal the value of NewReplicaCount plus 1 to account for the primary node. If this member of --replica-configuration is omitted, ElastiCache for Redis chooses the Availability Zone for each of the new replicas.

**Important**
You must include either the **--new-replica-count** or **--replica-configuration** parameter, but not both, in your call.

**Example**

The following example increases the number of replicas in the replication group sample-repl-group to three. When the example is finished, there are three replicas in each node group. This number applies whether this is a Redis (cluster mode disabled) group with a single node group or a Redis (cluster mode enabled) group with multiple node groups.

For Linux, macOS, or Unix:

```
aws elasticache increase-replica-count \
  --replication-group-id sample-repl-group \
  --new-replica-count 3 \
  --apply-immediately
```

For Windows:

```
aws elasticache increase-replica-count ^
  --replication-group-id sample-repl-group ^
  --new-replica-count 3 ^
  --apply-immediately
```

The following example increases the number of replicas in the replication group sample-repl-group to the value specified for the two specified node groups. Given that there are multiple node groups, this is a Redis (cluster mode enabled) replication group. When specifying the optional PreferredAvailabilityZones, the number of Availability Zones listed must equal the value of NewReplicaCount plus 1 more. This approach accounts for the primary node for the group identified by NodeGroupId.

For Linux, macOS, or Unix:

```
aws elasticache increase-replica-count \
  --replication-group-id sample-repl-group \
  --replica-configuration \
```
Changing the Number of Replicas

NodeGroupId=0001, NewReplicaCount=2, PreferredAvailabilityZones=us-east-1a, us-east-1b, us-east-1c, us-east-1d
NodeGroupId=0003, NewReplicaCount=3, PreferredAvailabilityZones=us-east-1a, us-east-1b, us-east-1c, us-east-1d
--apply-immediately

For Windows:

aws elasticache increase-replica-count ^
   --replication-group-id sample-repl-group ^
   --replica-configuration ^
       NodeGroupId=0001, NewReplicaCount=2, PreferredAvailabilityZones=us-east-1a, us-east-1b, us-east-1c, us-east-1d ^
       NodeGroupId=0003, NewReplicaCount=3, PreferredAvailabilityZones=us-east-1a, us-east-1b, us-east-1c, us-east-1d ^
   --apply-immediately

For more information about increasing the number of replicas using the CLI, see increase-replica-count in the Amazon ElastiCache Command Line Reference.

Using the ElastiCache API

To increase the number of replicas in a Redis shard, use the IncreaseReplicaCount action with the following parameters:

- ReplicationGroupId – Required. Identifies which replication group you want to increase the number of replicas in.
- ApplyImmediately – Required. Specifies whether to increase the replica count immediately (ApplyImmediately=True) or at the next maintenance window (ApplyImmediately=False). Currently, ApplyImmediately=False is not supported.
- NewReplicaCount – Optional. Specifies the number of replica nodes you want when finished, up to a maximum of five. Use this parameter for Redis (cluster mode disabled) replication groups where there is only one node group, or Redis (cluster mode enabled) groups where you want all node groups to have the same number of replicas. If this value is not larger than the current number of replicas in the node group, the call fails with an exception.
- ReplicaConfiguration – Optional. Allows you to set the number of replicas and Availability Zones for each node group independently. Use this parameter for Redis (cluster mode enabled) groups where you want to configure each node group independently.

ReplicaConfiguration has three optional members:

- NodeGroupId – The four-digit ID for the node group you are configuring. For Redis (cluster mode disabled) replication groups, the node group (shard) ID is always 0001. To find a Redis (cluster mode enabled) node group's (shard's) ID, see Finding a Shard's ID (p. 128).
- NewReplicaCount – The number of replicas that you want in this node group at the end of this operation. The value must be more than the current number of replicas and a maximum of five. If this value is not larger than the current number of replicas in the node group, the call fails with an exception.
- PreferredAvailabilityZones – A list of PreferredAvailabilityZone strings that specify which Availability Zones the replication group's nodes are to be in. The number of PreferredAvailabilityZone values must equal the value of NewReplicaCount plus 1 to account for the primary node. If this member of ReplicaConfiguration is omitted, ElastiCache for Redis chooses the Availability Zone for each of the new replicas.

Important
You must include either the NewReplicaCount or ReplicaConfiguration parameter, but not both, in your call.
Example

The following example increases the number of replicas in the replication group `sample-repl-group` to three. When the example is finished, there are three replicas in each node group. This number applies whether this is a Redis (cluster mode disabled) group with a single node group or a Redis (cluster mode enabled) group with multiple node groups.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
    ?Action=IncreaseReplicaCount
    &ApplyImmediately=True
    &NewReplicaCount=3
    &ReplicationGroupId=sample-repl-group
    &Version=2015-02-02
    &SignatureVersion=4
    &SignatureMethod=HmacSHA256
    &Timestamp=20150202T192317Z
    &X-Amz-Credential=<credential>
```

The following example increases the number of replicas in the replication group `sample-repl-group` to the value specified for the two specified node groups. Given that there are multiple node groups, this is a Redis (cluster mode enabled) replication group. When specifying the optional `PreferredAvailabilityZones`, the number of Availability Zones listed must equal the value of `NewReplicaCount` plus 1 more. This approach accounts for the primary node, for the group identified by `NodeGroupId`.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
    ?Action=IncreaseReplicaCount
    &ApplyImmediately=True
    &ReplicaConfiguration.ConfigureShard.1.NodeGroupId=0001
    &ReplicaConfiguration.ConfigureShard.1.NewReplicaCount=2
    &ReplicaConfiguration.ConfigureShard.1.PreferredAvailabilityZones.PreferredAvailabilityZone.1=us-east-1a
    &ReplicaConfiguration.ConfigureShard.1.PreferredAvailabilityZones.PreferredAvailabilityZone.2=us-east-1c
    &ReplicaConfiguration.ConfigureShard.1.PreferredAvailabilityZones.PreferredAvailabilityZone.3=us-east-1b
    &ReplicaConfiguration.ConfigureShard.2.NodeGroupId=0003
    &ReplicaConfiguration.ConfigureShard.2.NewReplicaCount=3
    &ReplicaConfiguration.ConfigureShard.2.PreferredAvailabilityZones.PreferredAvailabilityZone.1=us-east-1a
    &ReplicaConfiguration.ConfigureShard.2.PreferredAvailabilityZones.PreferredAvailabilityZone.2=us-east-1b
    &ReplicaConfiguration.ConfigureShard.2.PreferredAvailabilityZones.PreferredAvailabilityZone.3=us-east-1c
    &ReplicaConfiguration.ConfigureShard.2.PreferredAvailabilityZones.PreferredAvailabilityZone.4=us-east-1c
    &ReplicationGroupId=sample-repl-group
    &Version=2015-02-02
    &SignatureVersion=4
    &SignatureMethod=HmacSHA256
    &Timestamp=20150202T192317Z
    &X-Amz-Credential=<credential>
```

For more information about increasing the number of replicas using the API, see `IncreaseReplicaCount` in the Amazon ElastiCache API Reference.
Decreasing the Number of Replicas in a Shard

You can decrease the number of replicas in a shard for Redis (cluster mode enabled), or in a replication group for Redis (cluster mode disabled):

- For Redis (cluster mode disabled), you can decrease the number of replicas to one if Multi-AZ with Automatic Failover is enabled, and to zero if it isn't enabled.
- For Redis (cluster mode enabled), you can decrease the number of replicas to zero. However, you can’t fail over to a replica if your primary node fails.

You can use the AWS Management Console, the AWS CLI or the ElastiCache API to decrease the number of replicas in a node group (shard) or replication group.

Topics

- Using the AWS Management Console (p. 196)
- Using the AWS CLI (p. 196)
- Using the ElastiCache API (p. 198)

Using the AWS Management Console

The following procedure uses the console to decrease the number of replicas in a Redis (cluster mode enabled) replication group.

To decrease the number of replicas in a Redis shard

2. In the navigation pane, choose Redis, then choose the name of the replication group from which you want to delete replicas.
3. Choose the box for each shard you want to remove a replica node from.
4. Choose Delete replicas.
5. Complete the Delete Replicas from Shards page:
   a. For New number of replicas/shard, enter the number of replicas that you want the selected shards to have. This number must be greater than or equal to 1. We recommend at least two replicas per shard as a working minimum.
   b. Choose Delete to delete the replicas or Cancel to cancel the operation.

Important

- If you don’t specify the replica nodes to be deleted, ElastiCache for Redis automatically selects replica nodes for deletion. While doing so, ElastiCache for Redis attempts to retain the Multi-AZ architecture for your replication group followed by retaining replicas with minimum replication lag with the master.
- You can’t delete the primary or master nodes in a replication group. If you specify a primary node for deletion, the operation fails with an error event indicating that the primary node was selected for deletion.

Using the AWS CLI

To decrease the number of replicas in a Redis shard, use the decrease-replica-count command with the following parameters:
--replication-group-id – Required. Identifies which replication group you want to decrease the number of replicas in.

--apply-immediately or --no-apply-immediately – Required. Specifies whether to decrease the replica count immediately (--apply-immediately) or at the next maintenance window (--no-apply-immediately). Currently, --no-apply-immediately is not supported.

--new-replica-count – Optional. Specifies the number of replica nodes that you want. The value of --new-replica-count must be a valid value less than the current number of replicas in the node groups. For minimum permitted values, see Decreasing the Number of Replicas in a Shard (p. 196). If the value of --new-replica-count doesn't meet this requirement, the call fails.

--replicas-to-remove – Optional. Contains a list of node IDs specifying the replica nodes to remove.

--replica-configuration – Optional. Allows you to set the number of replicas and Availability Zones for each node group independently. Use this parameter for Redis (cluster mode enabled) groups where you want to configure each node group independently.

--replica-configuration has three optional members:

- NodeGroupId – The four-digit ID for the node group that you are configuring. For Redis (cluster mode disabled) replication groups, the shard ID is always 0001. To find a Redis (cluster mode enabled) node group's (shard's) ID, see Finding a Shard's ID (p. 128).

- NewReplicaCount – An optional parameter that specifies the number of replica nodes you want. The value of NewReplicaCount must be a valid value less than the current number of replicas in the node groups. For minimum permitted values, see Decreasing the Number of Replicas in a Shard (p. 196). If the value of NewReplicaCount doesn't meet this requirement, the call fails.

- PreferredAvailabilityZones – A list of PreferredAvailabilityZone strings that specify which Availability Zones the replication group's nodes are in. The number of PreferredAvailabilityZone values must equal the value of NewReplicaCount plus 1 to account for the primary node. If this member of --replica-configuration is omitted, ElastiCache for Redis chooses the Availability Zone for each of the new replicas.

**Important**
You must include one and only one of the --new-replica-count, --replicas-to-remove, or --replica-configuration parameters.

**Example**
The following example uses --new-replica-count to decrease the number of replicas in the replication group sample-repl-group to one. When the example is finished, there is one replica in each node group. This number applies whether this is a Redis (cluster mode disabled) group with a single node group or a Redis (cluster mode enabled) group with multiple node groups.

For Linux, macOS, or Unix:

```bash
aws elasticache decrease-replica-count
  --replication-group-id sample-repl-group \n  --new-replica-count 1 \n  --apply-immediately
```

For Windows:

```bash
aws elasticache decrease-replica-count ^
  --replication-group-id sample-repl-group ^
  --new-replica-count 1 ^
  --apply-immediately
```

The following example decreases the number of replicas in the replication group sample-repl-group by removing two specified replicas (0001 and 0003) from the node group.
For Linux, macOS, or Unix:

```bash
aws elasticache decrease-replica-count \
    --replication-group-id sample-repl-group \
    --replicas-to-remove 0001,0003 \
    --apply-immediately
```

For Windows:

```bash
aws elasticache decrease-replica-count ^
    --replication-group-id sample-repl-group ^
    --replicas-to-remove 0001,0003 ^
    --apply-immediately
```

The following example uses --replica-configuration to decrease the number of replicas in the replication group `sample-repl-group` to the value specified for the two specified node groups. Given that there are multiple node groups, this is a Redis (cluster mode enabled) replication group. When specifying the optional `PreferredAvailabilityZones`, the number of Availability Zones listed must equal the value of `NewReplicaCount` plus 1 more. This approach accounts for the primary node for the group identified by `NodeGroupId`.

For Linux, macOS, or Unix:

```bash
aws elasticache decrease-replica-count \
    --replication-group-id sample-repl-group \
    --replica-configuration \
        NodeGroupId=0001,NewReplicaCount=1,PreferredAvailabilityZones=us-east-1a,us-east-1c \
        NodeGroupId=0003,NewReplicaCount=2,PreferredAvailabilityZones=us-east-1a,us-east-1b,us-east-1c \
    --apply-immediately
```

For Windows:

```bash
aws elasticache decrease-replica-count ^
    --replication-group-id sample-repl-group ^
    --replica-configuration ^
        NodeGroupId=0001,NewReplicaCount=2,PreferredAvailabilityZones=us-east-1a,us-east-1c ^
        NodeGroupId=0003,NewReplicaCount=3,PreferredAvailabilityZones=us-east-1a,us-east-1b,us-east-1c ^
    --apply-immediately
```

For more information about decreasing the number of replicas using the CLI, see `decrease-replica-count` in the Amazon ElastiCache Command Line Reference.

**Using the ElastiCache API**

To decrease the number of replicas in a Redis shard, use the `DecreaseReplicaCount` action with the following parameters:

- **ReplicationGroupId** – Required. Identifies which replication group you want to decrease the number of replicas in.
- **ApplyImmediately** – Required. Specifies whether to decrease the replica count immediately (`ApplyImmediately=True`) or at the next maintenance window (`ApplyImmediately=False`). Currently, `ApplyImmediately=False` is not supported.
- **NewReplicaCount** – Optional. Specifies the number of replica nodes you want. The value of `NewReplicaCount` must be a valid value less than the current number of replicas in the node groups.
For minimum permitted values, see Decreasing the Number of Replicas in a Shard (p. 196). If the value of `--new-replica-count` doesn't meet this requirement, the call fails.

- **ReplicasToRemove** – Optional. Contains a list of node IDs specifying the replica nodes to remove.
- **ReplicaConfiguration** – Optional. Contains a list of node groups that allows you to set the number of replicas and Availability Zones for each node group independently. Use this parameter for Redis (cluster mode enabled) groups where you want to configure each node group independently.

**ReplicaConfiguration** has three optional members:

- **NodeGroupId** – The four-digit ID for the node group you are configuring. For Redis (cluster mode disabled) replication groups, the node group ID is always 0001. To find a Redis (cluster mode enabled) node group's (shard's) ID, see Finding a Shard's ID (p. 128).
- **NewReplicaCount** – The number of replicas that you want in this node group at the end of this operation. The value must be less than the current number of replicas down to a minimum of 1 if Multi-AZ with Automatic Failover is enabled or 0 if Multi-AZ with Automatic Failover isn't enabled. If this value is not less than the current number of replicas in the node group, the call fails with an exception.
- **PreferredAvailabilityZones** – A list of `PreferredAvailabilityZone` strings that specify which Availability Zones the replication group's nodes are in. The number of `PreferredAvailabilityZone` values must equal the value of `NewReplicaCount` plus 1 to account for the primary node. If this member of `ReplicaConfiguration` is omitted, ElastiCache for Redis chooses the Availability Zone for each of the new replicas.

**Important**

You must include one and only one of the `NewReplicaCount`, `ReplicasToRemove`, or `ReplicaConfiguration` parameters.

**Example**

The following example uses `NewReplicaCount` to decrease the number of replicas in the replication group `sample-repl-group` to one. When the example is finished, there is one replica in each node group. This number applies whether this is a Redis (cluster mode disabled) group with a single node group or a Redis (cluster mode enabled) group with multiple node groups.

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=DecreaseReplicaCount
  &ApplyImmediately=True
  &NewReplicaCount=1
  &ReplicationGroupId=sample-repl-group
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &X-Amz-Credential=<credential>
```

The following example decreases the number of replicas in the replication group `sample-repl-group` by removing two specified replicas (0001 and 0003) from the node group.

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=DecreaseReplicaCount
  &ApplyImmediately=True
  &ReplicasToRemove.ReplicaToRemove.1=0001
  &ReplicasToRemove.ReplicaToRemove.2=0003
  &ReplicationGroupId=sample-repl-group
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
```

**API Version 2015-02-02**

199
The following example uses ReplicaConfiguration to decrease the number of replicas in the replication group sample-repl-group to the value specified for the two specified node groups. Given that there are multiple node groups, this is a Redis (cluster mode enabled) replication group. When specifying the optional PreferredAvailabilityZones, the number of Availability Zones listed must equal the value of NewReplicaCount plus 1 more. This approach accounts for the primary node for the group identified by NodeGroupId.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DecreaseReplicaCount
 &ApplyImmediately=True
 &ReplicaConfiguration.ConfigureShard.1.NodeGroupId=0001
 &ReplicaConfiguration.ConfigureShard.1.NewReplicaCount=1
 &ReplicaConfiguration.ConfigureShard.1.PreferredAvailabilityZones_PreferredAvailabilityZone.1=us-east-1a
 &ReplicaConfiguration.ConfigureShard.1.PreferredAvailabilityZones_PreferredAvailabilityZone.2=us-east-1c
 &ReplicaConfiguration.ConfigureShard.2.NodeGroupId=0003
 &ReplicaConfiguration.ConfigureShard.2.NewReplicaCount=2
 &ReplicaConfiguration.ConfigureShard.2.PreferredAvailabilityZones_PreferredAvailabilityZone.1=us-east-1a
 &ReplicaConfiguration.ConfigureShard.2.PreferredAvailabilityZones_PreferredAvailabilityZone.2=us-east-1b
 &ReplicaConfiguration.ConfigureShard.2.PreferredAvailabilityZones_PreferredAvailabilityZone.4=us-east-1c
 &ReplicationGroupId=sample-repl-group
 &Version=2015-02-02
 &SignatureVersion=4
 &SignatureMethod=HmacSHA256
 &Timestamp=20150202T192317Z
 &X-Amz-Credential=<credential>
```

For more information about decreasing the number of replicas using the API, see DecreaseReplicaCount in the Amazon ElastiCache API Reference.

### Adding a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups

Information in the following topic applies to Redis (cluster mode disabled) replication groups only.

As your read traffic increases, you might want to spread those reads across more nodes and reduce the read pressure on any one node. In this topic, you can find how to add a read replica to a Redis (cluster mode disabled) cluster.

A Redis (cluster mode disabled) replication group can have a maximum of five read replicas. If you attempt to add a read replica to a replication group that already has five read replicas, the operation fails.

For information about adding replicas to a Redis (cluster mode enabled) replication group, see the following:

- Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)
- Increasing the Number of Replicas in a Shard (p. 192)
You can add a read replica to a Redis (cluster mode disabled) cluster using the ElastiCache Console, the AWS CLI, or the ElastiCache API.

**Related topics**

- Adding Nodes to a Cluster (p. 104)
- Adding a Read Replica to a Replication Group (AWS CLI) (p. 201)
- Adding a Read Replica to a Replication Group Using the API (p. 201)

### Adding a Read Replica to a Replication Group (AWS CLI)

To add a read replica to a Redis (cluster mode disabled) replication group, use the AWS CLI `create-cache-cluster` command, with the parameter `--replication-group-id` to specify which replication group to add the cluster (node) to.

The following example creates the cluster `my-read replica` and adds it to the replication group `my-replication-group`. The node types, parameter groups, security groups, maintenance window, and other settings for the read replica are the same as for the other nodes in `my-replication-group`.

For Linux, macOS, or Unix:

```bash
aws elasticache create-cache-cluster \
    --cache-cluster-id my-read replica \
    --replication-group-id my-replication-group
```

For Windows:

```bash
aws elasticache create-cache-cluster ^
    --cache-cluster-id my-read replica ^
    --replication-group-id my-replication-group
```

For more information on adding a read replica using the CLI, see `create-cache-cluster` in the Amazon ElastiCache Command Line Reference.

### Adding a Read Replica to a Replication Group Using the API

To add a read replica to a Redis (cluster mode disabled) replication group, use the ElastiCache `CreateCacheCluster` operation, with the parameter `ReplicationGroupId` to specify which replication group to add the cluster (node) to.

The following example creates the cluster `myReadReplica` and adds it to the replication group `myReplicationGroup`. The node types, parameter groups, security groups, maintenance window, and other settings for the read replica are the same as for the other nodes in `myReplicationGroup`.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=CreateCacheCluster
&CacheClusterId=myReadReplica
&ReplicationGroupId=myReplicationGroup
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>
```

For more information on adding a read replica using the API, see `CreateCacheCluster` in the Amazon ElastiCache API Reference.
Deleting a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups

Information in the following topic applies to Redis (cluster mode disabled) replication groups only.

As read traffic on your Redis replication group changes, you might want to add or remove read replicas. Removing a node from a Redis (cluster mode disabled) replication group is the same as just deleting a cluster, though there are restrictions:

- You cannot remove the primary from a replication group. If you want to delete the primary, do the following:
  1. Promote a read replica to primary. For more information on promoting a read replica to primary, see Promoting a Read Replica to Primary, for Redis (cluster mode disabled) Replication Groups (p. 203).
  2. Delete the old primary. For a restriction on this method, see the next point.
- If Multi-AZ is enabled on a replication group, you can't remove the last read replica from the replication group. In this case, do the following:
  1. Modify the replication group by disabling Multi-AZ. For more information, see Modifying a Replication Group (p. 188).
  2. Delete the read replica.

You can remove a read replica from a Redis (cluster mode disabled) replication group using the ElastiCache console, the AWS CLI for ElastiCache, or the ElastiCache API.

For directions on deleting a cluster from a Redis replication group, see the following:

- Using the AWS Management Console (p. 117)
- Using the AWS CLI (p. 117)
- Using the ElastiCache API (p. 118)
- Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)
- Decreasing the Number of Replicas in a Shard (p. 196)
Promoting a Read Replica to Primary, for Redis (cluster mode disabled) Replication Groups

Information in the following topic applies to only Redis (cluster mode disabled) replication groups.

You can promote a Redis (cluster mode disabled) read replica to primary using the AWS Management Console, the AWS CLI, or the ElastiCache API. You can't promote a read replica to primary while Multi-AZ with Automatic Failover is enabled on the Redis (cluster mode disabled) replication group. To promote a Redis (cluster mode disabled) replica to primary on a Multi-AZ enabled replication group, do the following:

1. Modify the replication group to disable Multi-AZ (doing this doesn’t require that all your clusters be in the same Availability Zone). For more information, see Modifying a Replication Group (p. 188).
2. Promote the read replica to primary.
3. Modify the replication group to re-enable Multi-AZ.

Multi-AZ with Automatic Failover is not available on replication groups running Redis 2.6.13 or earlier.

Using the AWS Management Console

The following procedure uses the console to promote a replica node to primary.

To promote a read replica to primary (console)

2. If the replica you want to promote is a member of a Redis (cluster mode disabled) replication group where Multi-AZ is enabled, modify the replication group to disable Multi-AZ before you proceed. For more information, see Modifying a Replication Group (p. 188).
3. Choose Redis, then from the list of clusters, choose the replication group that you want to modify. This replication group must be running the "Redis" engine, not the "Clustered Redis" engine, and must have two or more nodes.
4. From the list of nodes, choose the replica node you want to promote to primary, then for Actions, choose Promote.
5. In the Promote Read Replica dialog box, do the following:
   a. For Apply Immediately, choose Yes to promote the read replica immediately, or No to promote it at the cluster's next maintenance window.
   b. Choose Promote to promote the read replica or Cancel to cancel the operation.
6. If the cluster was Multi-AZ enabled before you began the promotion process, wait until the replication group's status is available, then modify the cluster to re-enable Multi-AZ. For more information, see Modifying a Replication Group (p. 188).

Using the AWS CLI

You can't promote a read replica to primary if the replication group is Multi-AZ enabled. In some cases, the replica that you want to promote might be a member of a replication group where Multi-AZ is enabled. In these cases, you must modify the replication group to disable Multi-AZ before you proceed. Doing this doesn't require that all your clusters be in the same Availability Zone. For more information on modifying a replication group, see Modifying a Replication Group (p. 188).

The following AWS CLI command modifies the replication group sample-repl-group, making the read replica my-replica-1 the primary in the replication group.
Finding Connection Endpoints

For Linux, macOS, or Unix:

```
aws elasticache modify-replication-group \
   --replication-group-id sample-repl-group \
   --primary-cluster-id my-replica-1
```

For Windows:

```
aws elasticache modify-replication-group ^
   --replication-group-id sample-repl-group ^
   --primary-cluster-id my-replica-1
```

For more information on modifying a replication group, see modify-replication-group in the Amazon ElastiCache Command Line Reference.

Using the ElastiCache API

You can't promote a read replica to primary if the replication group is Multi-AZ enabled. In some cases, the replica that you want to promote might be a member of a replication group where Multi-AZ is enabled. In these cases, you must modify the replication group to disable Multi-AZ before you proceed. Doing this doesn't require that all your clusters be in the same Availability Zone. For more information on modifying a replication group, see Modifying a Replication Group (p. 188).

The following ElastiCache API action modifies the replication group myReplGroup, making the read replica myReplica-1 the primary in the replication group.

```
https://elasticache.us-west-2.amazonaws.com/
   ?Action=ModifyReplicationGroup
   &ReplicationGroupId=myReplGroup
   &PrimaryClusterId=myReplica-1
   &Version=2014-12-01
   &SignatureVersion=4
   &SignatureMethod=HmacSHA256
   &X-Amz-Algorithm=AWS4-HMAC-SHA256
   &X-Amz-Date=20141201T220302Z
   &X-Amz-Expires=20141201T220302Z
   &X-Amz-Credential=<credential>
   &X-Amz-Signature=<signature>
```

For more information on modifying a replication group, see ModifyReplicationGroup in the Amazon ElastiCache API Reference.

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

- **Redis standalone node**, use the node's endpoint for both read and write operations.

- **Redis (cluster mode disabled) clusters**, use the Primary Endpoint for all write operations. Use the Reader Endpoint to evenly split incoming connections to the endpoint between all read replicas.
Use the individual *Node Endpoints* for read operations (In the API/CLI these are referred to as Read Endpoints).

- **Redis (cluster mode enabled) clusters**, use the cluster's *Configuration Endpoint* for all operations. You must use a client that supports Redis Cluster (Redis 3.2). You can still read from individual node endpoints (In the API/CLI these are referred to as Read Endpoints).

The following sections guide you through discovering the endpoints you'll need for the engine you're running.

**Topics**

- Finding a Redis (Cluster Mode Disabled) Cluster's Endpoints (Console) (p. 206)
- Finding Endpoints for a Redis (Cluster Mode Enabled) Cluster (Console) (p. 208)
- Finding Endpoints (AWS CLI) (p. 210)
- Finding Endpoints (ElastiCache API) (p. 213)
Finding a Redis (Cluster Mode Disabled) Cluster's Endpoints (Console)

If a Redis (cluster mode disabled) cluster has only one node, the node's endpoint is used for both reads and writes. If a Redis (cluster mode disabled) cluster has multiple nodes, there are three types of endpoints; the primary endpoint, the reader endpoint and the node endpoints.

The primary endpoint is a DNS name that always resolves to the primary node in the cluster. The primary endpoint is immune to changes to your cluster, such as promoting a read replica to the primary role. For write activity, we recommend that your applications connect to the primary endpoint instead of connecting directly to the primary.

A reader endpoint will evenly split incoming connections to the endpoint between all read replicas in a ElastiCache for Redis cluster. Reader endpoints keep up with cluster changes in real-time as replicas are added or removed. You can place your ElastiCache for Redis cluster's multiple read replicas in different AWS Availability Zones (AZ) to ensure high availability of reader endpoints.

For read activity, applications can also connect to any node in the cluster. Unlike the primary endpoint, node endpoints resolve to specific endpoints. If you make a change in your cluster, such as adding or deleting a replica, you must update the node endpoints in your application.

To find a Redis (cluster mode disabled) cluster's endpoints

2. From the navigation pane, choose Redis.
   The clusters screen will appear with a list of Redis (cluster mode disabled) and Redis (cluster mode enabled) clusters.
3. To find the cluster's Primary and/or Reader endpoints, choose the box to the left of cluster's name.

   **Primary and Reader endpoints for a Redis (cluster mode disabled) cluster**

   If there is only one node in the cluster, there is no primary endpoint and you can continue at the next step.
4. If the Redis (cluster mode disabled) cluster has replica nodes, you can find the cluster's replica node endpoints by choosing the cluster's name.
   The nodes screen appears with each node in the cluster, primary and replicas, listed with its endpoint.
Node endpoints for a Redis (cluster mode disabled) cluster

5. To copy an endpoint to your clipboard:
   a. One endpoint at a time, find then highlight the endpoint you want to copy.
   b. Right-click the highlighted endpoint, then choose Copy from the context menu.

The highlighted endpoint is now copied to your clipboard.

A Redis (cluster mode disabled) primary endpoint looks something like the following. There is a difference depending upon whether or not In-Transit encryption is enabled.

In-transit encryption not enabled

```
clusterName.xxxxxx.nodeId.regionAndAz.cache.amazonaws.com:port
redis-01.7abc2d.0001.usw2.cache.amazonaws.com:6379
```

In-transit encryption enabled

```
master.clusterName.xxxxxx.regionAndAz.cache.amazonaws.com:port
master.ncit.ameaqx.use1.cache.amazonaws.com:6379
```
Finding Endpoints for a Redis (Cluster Mode Enabled) Cluster (Console)

Use the Configuration Endpoint for both read and write operations. Redis determines which of the cluster's node to access.

The following procedure demonstrates how to find and copy Redis (cluster mode enabled) cluster endpoints.

To find the configuration endpoint for a Redis (cluster mode enabled) cluster

2. From the navigation pane, choose Redis.
   
   A list of clusters running any version of Redis appears.
3. From the list of clusters, choose the box to the left of a cluster running "Clustered Redis".
   
   The screen expands showing details about the selected cluster.
4. Locate the Configuration endpoint.

Configuration endpoint for a Redis (cluster mode enabled) cluster

To find the node endpoints for a Redis (cluster mode enabled) cluster

2. From the navigation pane, choose Redis.
   
   A list of clusters running any version of Redis appears.
3. From the list of clusters, choose the cluster name of a cluster running "Clustered Redis".
   
   The shards page opens.
4. Choose the name of the shard you want node endpoint for.
   
   A list of the shard's nodes appears with each node's endpoint.
5. Locate the Endpoint column and read the endpoint for each node.
To copy an endpoint to your clipboard

1. Find the endpoint you want to copy using one of the preceding procedures.
2. Highlight the endpoint that you want to copy.
3. Right-click the highlighted endpoint and choose Copy from the context menu.

The highlighted endpoint is now copied to your clipboard.

A Redis (cluster mode enabled) configuration endpoint looks something like the following.

**In-transit encryption not enabled**

```
clusterName.xxxxxx.regionAndAz.cache.amazonaws.com:port
rce.ameaqx.use1.cache.amazonaws.com:6379
```

**In-transit encryption enabled**

```
clustercfg.clusterName.xxxxxx.regionAndAz.cache.amazonaws.com:port
clustercfg.rce.ameaqx.use1.cache.amazonaws.com:6379
```
Finding Endpoints (AWS CLI)

You can use the AWS CLI for Amazon ElastiCache to discover the endpoints for nodes, clusters, and replication groups.

Topics

- Finding Endpoints for Nodes and Clusters (AWS CLI) (p. 210)
- Finding the Endpoints for Replication Groups (AWS CLI) (p. 211)

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 Redis clusters, the command returns the cluster 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 cluster information for the single-node Redis (cluster mode disabled) cluster mycluster.

Important

The parameter --cache-cluster-id can be used with single-node Redis (cluster mode disabled) cluster id or specific node ids in Redis replication groups. The --cache-cluster-id of a Redis replication group is a 4-digit value such as 0001. If --cache-cluster-id is the id of a cluster (node) in a Redis replication group, the replication-group-id is included in the output.

For Linux, macOS, or Unix:

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

For Windows:

```
aws elasticache describe-cache-clusters ^
   --cache-cluster-id redis-cluster ^
   --show-cache-node-info
```

Output from the above operation should look something like this (JSON format).

```
{
   "CacheClusters": [
      {
         "CacheClusterStatus": "available",
         "SecurityGroups": [
            {
               "SecurityGroupId": "sg-77186e0d",
               "Status": "active"
            }
         ],
         "CacheNodes": [
            {
               "CustomerAvailabilityZone": "us-east-1b",
               "CacheNodeStatus": "available",
```
For more information, see the topic describe-cache-clusters.

Finding the Endpoints for Replication Groups (AWS CLI)

You can use the AWS CLI to discover the endpoints for a replication group and its clusters with the describe-replication-groups command. The command returns the replication group's primary endpoint and a list of all the clusters (nodes) in the replication group with their endpoints, along with the reader endpoint.

The following operation retrieves the primary endpoint and reader endpoint for the replication group myreplgroup. Use the primary endpoint for all write operations.

```
aws elasticache describe-replication-groups \
  --replication-group-id myreplgroup
```

For Windows:

```
aws elasticache describe-replication-groups ^ \
  --replication-group-id myreplgroup
```

Output from this operation should look something like this (JSON format).

```json
{
  "ReplicationGroups": [
    {
      "Status": "available",
```
"Description": "test",
"NodeGroups": [
  {
    "Status": "available",
    "NodeGroupMembers": [
      {
        "CurrentRole": "primary",
        "PreferredAvailabilityZone": "us-west-2a",
        "CacheNodeId": "0001",
        "ReadEndpoint": {
          "Port": 6379,
          "Address": "myreplgroup-001.1abc4d.0001.usw2.cache.amazonaws.com"
        },
        "CacheClusterId": "myreplgroup-001"
      },
      {
        "CurrentRole": "replica",
        "PreferredAvailabilityZone": "us-west-2b",
        "CacheNodeId": "0001",
        "ReadEndpoint": {
          "Port": 6379,
          "Address": "myreplgroup-002.1abc4d.0001.usw2.cache.amazonaws.com"
        },
        "CacheClusterId": "myreplgroup-002"
      },
      {
        "CurrentRole": "replica",
        "PreferredAvailabilityZone": "us-west-2c",
        "CacheNodeId": "0001",
        "ReadEndpoint": {
          "Port": 6379,
          "Address": "myreplgroup-003.1abc4d.0001.usw2.cache.amazonaws.com"
        },
        "CacheClusterId": "myreplgroup-003"
      }
    ],
    "NodeGroupId": "0001",
    "PrimaryEndpoint": {
      "Port": 6379,
      "Address": "myreplgroup.1abc4d.ng.0001.usw2.cache.amazonaws.com"
    },
    "ReaderEndpoint": {
      "Port": 6379,
      "Address": "myreplgroup-ro.1abc4d.ng.0001.usw2.cache.amazonaws.com"
    }
  }
],
"ReplicationGroupId": "myreplgroup",
"AutomaticFailover": "enabled",
"SnapshottingClusterId": "myreplgroup-002",
"MemberClusters": [
  "myreplgroup-001",
  "myreplgroup-002",
  "myreplgroup-003"
],
"PendingModifiedValues": {}]
}

For more information, see describe-replication-groups in the AWS CLI Command Reference.
Finding Endpoints (ElastiCache API)

You can use the Amazon ElastiCache API to discover the endpoints for nodes, clusters, and replication groups.

Topics
- Finding Endpoints for Nodes and Clusters (ElastiCache API) (p. 213)
- Finding Endpoints for Replication Groups (ElastiCache API) (p. 213)

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 Redis clusters, the command returns the cluster endpoint. If you include the optional parameter ShowCacheNodeInfo, the action also returns the endpoints of the individual nodes in the cluster.

Example


Finding Endpoints for Replication Groups (ElastiCache API)

You can use the ElastiCache API to discover the endpoints for a replication group and its clusters with the DescribeReplicationGroups action. The action returns the replication group's primary endpoint and a list of all the clusters in the replication group with their endpoints, along with the reader endpoint.

The following operation retrieves the primary endpoint (PrimaryEndpoint), reader endpoint (ReaderEndpoint) and individual node endpoints (ReadEndpoint) for the replication group myreplgroup. Use the primary endpoint for all write operations.


For more information, see DescribeReplicationGroups.

Backup and Restore for ElastiCache for Redis

Amazon ElastiCache clusters running Redis can back up their data. You can use the backup to restore a cluster or seed a new cluster. The backup consists of the cluster’s metadata, along with all of the data in the cluster. All backups are written to Amazon Simple Storage Service (Amazon S3), which provides
durable storage. At any time, you can restore your data by creating a new Redis cluster and populating it with data from a backup. With ElastiCache, you can manage backups using the AWS Management Console, the AWS Command Line Interface (AWS CLI), and the ElastiCache API.

Beginning with Redis version 2.8.22, the backup method is selected based upon available memory. If there is sufficient available memory, a child process is spawned that writes all changes to the cache's reserved memory while the cache is being backed up. Depending on the number of writes to the cache during the backup process, this child process can consume all reserved memory, causing the backup to fail.

If there is insufficient memory available, a forkless, cooperative background process is employed. The forkless method can affect both latency and throughput. For more information, see How Synchronization and Backup are Implemented (p. 161).

For more information about the performance impact of the backup process, see Performance Impact of Backups (p. 215).

Following, you can find an overview of working with backup and restore.

**Important**

Though it's rare, sometimes the backup process fails to create a backup, including final backups. Insufficient reserved memory is often the cause of backup failures. Therefore, make sure that you have sufficient reserved memory before attempting a backup. If you have insufficient memory, you can either evict some keys or increase the value of `reserved-memory-percent`. For more information, see the following:

- Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464)
- Managing Reserved Memory (p. 466)

If you plan to delete cluster and it's important to preserve the data, you can take an extra precaution. To do this, create a manual backup first, verify that its status is `available`, and then delete the cluster. Doing this makes sure that if the backup fails, you still have the cluster data available. You can retry making a backup, following the best practices outlined preceding.

**Topics**

- Backup Constraints (p. 214)
- Backup Costs (p. 215)
- Performance Impact of Backups (p. 215)
- Scheduling Automatic Backups (p. 217)
- Making Manual Backups (p. 218)
- Creating a Final Backup (p. 224)
- Describing Backups (p. 227)
- Copying a Backup (p. 229)
- Exporting a Backup (p. 231)
- Restoring From a Backup with Optional Cluster Resizing (p. 239)
- Seeding a New Cluster with an Externally Created Backup (p. 242)
- Tagging Backups (p. 248)
- Deleting a Backup (p. 249)
- Append Only Files (AOF) in ElastiCache for Redis (p. 250)

**Backup Constraints**

Consider the following constraints when planning or making backups:
• At this time, backup and restore are supported only for clusters running on Redis.
• For Redis (cluster mode disabled) clusters, backup and restore aren't supported on cache.t1.micro nodes. All other cache node types are supported.
• For Redis (cluster mode enabled) clusters, backup and restore are supported for all node types.
• During any contiguous 24-hour period, you can create no more than 20 manual backups per node in the cluster.
• Redis (cluster mode enabled) only supports taking backups on the cluster level (for the API or CLI, the replication group level). Redis (cluster mode enabled) doesn't support taking backups at the shard level (for the API or CLI, the node group level).
• During the backup process, you can't run any other API or CLI operations on the cluster.

Backup Costs

Using ElastiCache, you can store one backup for each active Redis cluster free of charge. Storage space for additional backups is charged at a rate of $0.085/GB per month for all AWS Regions. There are no data transfer fees for creating a backup, or for restoring data from a backup to a Redis cluster.

Performance Impact of Backups

The backup process depends upon which Redis version you're running. Beginning with Redis 2.8.22, the process is forkless.

Backups When Running Redis 2.8.22 and Later

In versions 2.8.22 and later, Redis backups choose between two methods. If there isn't enough memory to support a forked backup, ElastiCache use a forkless method that uses cooperative background processing. If there is enough memory to support a forked save process, the same process is used as in earlier Redis versions.

If the write load is high during a forkless backup, writes to the cache are delayed. This delay makes sure that you don't accumulate too many changes and thus prevent a successful backup.

Backups When Running Redis Versions before 2.8.22

Backups are created using Redis' native BGSAVE operation. The Redis process on the cache node spawns a child process to write all the data from the cache to a Redis .rdb file. It can take up to 10 seconds to spawn the child process. During this time, the parent process is unable to accept incoming application requests. After the child process is running independently, the parent process resumes normal operations. The child process exits when the backup operation is complete.

While the backup is being written, additional cache node memory is used for new writes. If this additional memory usage exceeds the node's available memory, processing can become slow due to excessive paging, or fail.

Improving Backup Performance

The following are guidelines for improving backup performance.

• Set the reserved-memory-percent parameter – To mitigate excessive paging, we recommend that you set the reserved-memory-percent parameter. This parameter prevents Redis from consuming all of the node's available memory, and can help reduce the amount of paging. You might also see performance improvements by simply using a larger node. For more information about the reserved-memory and reserved-memory-percent parameters, see Managing Reserved Memory (p. 466).
Create backups from a read replica – If you are running Redis in a node group with more than one node, you can take a backup from the primary node or one of the read replicas. Because of the system resources required during BGSAVE, we recommend that you create backups from one of the read replicas. While the backup is being created from the replica, the primary node remains unaffected by BGSAVE resource requirements. The primary node can continue serving requests without slowing down.

If you delete a replication group and request a final backup, ElastiCache always takes the backup from the primary node. This ensures that you capture the very latest Redis data, before the replication group is deleted.
Scheduling Automatic Backups

For any Redis cluster, you can enable automatic backups. When automatic backups are enabled, ElastiCache creates a backup of the cluster on a daily basis. Automatic backups can help guard against data loss. In the event of a failure, you can create a new cluster, restoring your data from the most recent backup. The result is a warm-started cluster, preloaded with your data and ready for use. For more information, see Restoring From a Backup with Optional Cluster Resizing (p. 239).

When you schedule automatic backups, you should plan the following settings:

- **Backup window** – A period during each day when ElastiCache begins creating a backup. The minimum length for the backup window is 60 minutes. You can set the backup window for any time when it’s most convenient for you, or for a time of day that avoids doing backups during particularly high-utilization periods.

  If you don't specify a backup window, ElastiCache assigns one automatically.

- **Backup retention limit** – The number of days the backup is retained in Amazon S3. For example, if you set the retention limit to 5, then a backup taken today is retained for 5 days. When the retention limit expires, the backup is automatically deleted.

  The maximum backup retention limit is 35 days. If the backup retention limit is set to 0, automatic backups are disabled for the cluster.

You can enable or disable automatic backups on an existing Redis cluster or replication group by modifying it using the ElastiCache console, the AWS CLI, or the ElastiCache API. For more information on how to enable or disable automatic backups on an existing cluster or replication group, see Modifying an ElastiCache Cluster (p. 100) or Modifying a Replication Group (p. 188).

You can enable or disable automatic backups when creating a Redis cluster or replication group using the ElastiCache console, the AWS CLI, or the ElastiCache API. You can enable automatic backups when you create a Redis cluster by checking the Enable Automatic Backups box in the Advanced Redis Settings section. For more information, see step 2 of Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79). You can enable automatic backups when you create a Redis replication group if you are not using an existing cluster as the primary cluster. For more information, see Creating a Redis Replication Group from Scratch (p. 168).
Making Manual Backups

In addition to automatic backups, you can create a manual backup at any time. Unlike automatic backups, which are automatically deleted after a specified retention period, manual backups do not have a retention period after which they are automatically deleted. You must manually delete any manual backup. Even if you delete a cluster or node, any manual backups from that cluster or node are retained. If you no longer want to keep a manual backup, you must explicitly delete it yourself.

Manual backups are useful for testing and archiving. For example, suppose that you've developed a set of baseline data for testing purposes. You can create a manual backup of the data and restore it whenever you want. After you test an application that modifies the data, you can reset the data by creating a new cluster and restoring from your baseline backup. When the cluster is ready, you can test your applications against the baseline data again—and repeat this process as often as needed.

In addition to directly creating a manual backup, you can create a manual backup in one of the following ways:

- **Copying a Backup (p. 229)** It does not matter whether the source backup was created automatically or manually.
- **Creating a Final Backup (p. 224)** Create a backup immediately before deleting a cluster or node.

Other topics of import

- [Backup Constraints (p. 214)]()
- [Backup Costs (p. 215)]()
- [Performance Impact of Backups (p. 215)]()

You can create a manual backup of a node using the AWS Management Console, the AWS CLI, or the ElastiCache API.

Creating a Manual Backup (Console)

To create a backup of a cluster (console)

2. From the navigation pane, choose Redis. The Redis clusters screen appears.
3. Choose the box to the left of the name of the Redis cluster you want to back up.
4. Choose Backup.
5. In the Create Backup dialog, type in a name for your backup in the Backup Name box. We recommend that the name indicate which cluster was backed up and the date and time the backup was made.

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.
6. Choose Create Backup.
The status of the cluster changes to *snapshotting*. When the status returns to *available* the backup is complete.

## Creating a Manual Backup (AWS CLI)

To create a manual backup of a cluster using the AWS CLI, use the `create-snapshot` AWS CLI operation with the following parameters:

- `--cache-cluster-id`
  - If the cluster you’re backing up has no replica nodes, `--cache-cluster-id` is the name of the cluster you are backing up, for example `mycluster`.
  - If the cluster you’re backing up has one or more replica nodes, `--cache-cluster-id` is the name of the node in the cluster that you want to use for the backup. For example, the name might be `mycluster-002`.

  Use this parameter only when backing up a Redis (cluster mode disabled) cluster.

- `--replication-group-id` - Name of the Redis (cluster mode enabled) cluster (CLI/API: a replication group) to use as the source for the backup. Use this parameter when backing up a Redis (cluster mode enabled) cluster.

- `--snapshot-name` - Name of the snapshot to be created.

  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.

### Example 1: Backing Up a Redis (Cluster Mode Disabled) Cluster That Has No Replica Nodes

The following AWS CLI operation creates the backup `bkup-20150515` from the Redis (cluster mode disabled) cluster `myNonClusteredRedis` that has no read replicas.

For Linux, macOS, or Unix:

```
aws elasticache create-snapshot \
  --cache-cluster-id myNonClusteredRedis \
  --snapshot-name bkup-20150515
```

For Windows:

```
aws elasticache create-snapshot ^
  --cache-cluster-id myNonClusteredRedis ^
  --snapshot-name bkup-20150515
```
Example 2: Backing Up a Redis (Cluster Mode Disabled) Cluster with Replica Nodes

The following AWS CLI operation creates the backup bkup-20150515 from the Redis (cluster mode disabled) cluster myNonClusteredRedis. This backup has one or more read replicas.

For Linux, macOS, or Unix:

```bash
aws elasticache create-snapshot
   --cache-cluster-id myNonClusteredRedis-001
   --snapshot-name bkup-20150515
```

For Windows:

```bash
aws elasticache create-snapshot
   --cache-cluster-id myNonClusteredRedis-001
   --snapshot-name bkup-20150515
```

Example Output: Backing Up a Redis (Cluster Mode Disabled) Cluster with Replica Nodes

Output from the operation looks something like the following.

```
{
   "Snapshot": {
      "Engine": "redis",
      "CacheParameterGroupName": "default.redis3.2",
      "VpcId": "vpc-91280df6",
      "CacheClusterId": "myNonClusteredRedis-001",
      "SnapshotRetentionLimit": 0,
      "NumCacheNodes": 1,
      "SnapshotName": "bkup-20150515",
      "CacheClusterCreateTime": "2017-01-12T18:59:48.048Z",
      "AutoMinorVersionUpgrade": true,
      "PreferredAvailabilityZone": "us-east-1c",
      "SnapshotStatus": "creating",
      "SnapshotSource": "manual",
      "SnapshotWindow": "08:30-09:30",
      "EngineVersion": "3.2.4",
      "NodeSnapshots": [
         {
            "CacheSize": "",
            "CacheNodeId": "0001",
            "CacheNodeCreateTime": "2017-01-12T18:59:48.048Z"
         }
      ],
      "CacheSubnetGroupName": "default",
      "Port": 6379,
      "PreferredMaintenanceWindow": "wed:07:30-wed:08:30",
      "CacheNodeType": "cache.m3.2xlarge"
   }
}
```

Example 3: Backing Up a Cluster for Redis (Cluster Mode Enabled)

The following AWS CLI operation creates the backup bkup-20150515 from the Redis (cluster mode enabled) cluster myClusteredRedis. Note the use of `--replication-group-id` instead of `--cache-cluster-id` to identify the source.

For Linux, macOS, or Unix:

```
aws elasticache create-snapshot
   --replication-group-id myClusteredRedis-001
   --snapshot-name bkup-20150515
```

API Version 2015-02-02
aws elasticache create-snapshot \
  --replication-group-id myClusteredRedis \
  --snapshot-name bkup-20150515

For Windows:

aws elasticache create-snapshot ^
  --replication-group-id myClusteredRedis ^
  --snapshot-name bkup-20150515

Example Output: Backing Up a Redis (Cluster Mode Enabled) Cluster

Output from this operation looks something like the following.

```
{
  "Snapshot": {
    "Engine": "redis",
    "CacheParameterGroupName": "default.redis3.2.cluster.on",
    "VpcId": "vpc-91280df6",
    "NodeSnapshots": [
      {
        "CacheSize": "",
        "NodeGroupId": "0001"
      },
      {
        "CacheSize": "",
        "NodeGroupId": "0002"
      }
    ],
    "NumNodeGroups": 2,
    "SnapshotName": "bkup-20150515",
    "ReplicationGroupId": "myClusteredRedis",
    "AutoMinorVersionUpgrade": true,
    "SnapshotRetentionLimit": 1,
    "AutomaticFailover": "enabled",
    "SnapshotStatus": "creating",
    "SnapshotSource": "manual",
    "SnapshotWindow": "10:00-11:00",
    "EngineVersion": "3.2.4",
    "CacheSubnetGroupName": "default",
    "ReplicationGroupDescription": "2 shards 2 nodes each",
    "Port": 6379,
    "PreferredMaintenanceWindow": "sat:03:30-sat:04:30",
    "CacheNodeType": "cache.r3.large"
  }
}
```

Related Topics

For more information, see create-snapshot in the AWS CLI Command Reference.

Creating a Manual Backup (ElastiCache API)

To create a manual backup of a cluster using the ElastiCache API, use the CreateSnapshot ElastiCache API operation with the following parameters:

- **CacheClusterId**
  - If the cluster you're backing up has no replica nodes, *CacheClusterId* is the name of the cluster you are backing up, for example *mycluster*.  

API Version 2015-02-02

221
• If the cluster you're backing up has one or more replica nodes, CacheClusterId is the name of the node in the cluster that you want to use for the backup, for example mycluster-002.

Only use this parameter when backing up a Redis (cluster mode disabled) cluster.

• ReplicationGroupId – Name of the Redis (cluster mode enabled) cluster (CLI/API: a replication group) to use as the source for the backup. Use this parameter when backing up a Redis (cluster mode enabled) cluster.

• SnapshotName – Name of the snapshot to be created.

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.

API Code Examples

Example 1: Backing Up a Redis (Cluster Mode Disabled) Cluster That Has No Replica Nodes (p. 222)

Example 2: Backing Up a Redis (Cluster Mode Disabled) Cluster with Replica Nodes (p. 222)

Example 3: Backing Up a Redis (Cluster Mode Enabled) Cluster (p. 223)

Related Topics (p. 223)

Example 1: Backing Up a Redis (Cluster Mode Disabled) Cluster That Has No Replica Nodes

The following ElastiCache API operation creates the backup bkup-20150515 from the Redis (cluster mode disabled) cluster myNonClusteredRedis that has no read replicas.

https://elasticache.us-west-2.amazonaws.com/?Action=CreateSnapshot
&CacheClusterId=myNonClusteredRedis
&SnapshotName=bkup-20150515
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>

Example 2: Backing Up a Redis (Cluster Mode Disabled) Cluster with Replica Nodes

The following ElastiCache API operation creates the backup bkup-20150515 from the Redis (cluster mode disabled) cluster myNonClusteredRedis which has one or more read replicas.

https://elasticache.us-west-2.amazonaws.com/?Action=CreateSnapshot
&CacheClusterId=myNonClusteredRedis-001
&SnapshotName=bkup-20150515
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256

API Version 2015-02-02
222
Example 3: Backing Up a Redis (Cluster Mode Enabled) Cluster

The following ElastiCache API operation creates the backup bkup-20150515 from the Redis (cluster mode enabled) cluster myClusteredRedis. Note the use of ReplicationGroupId instead of CacheClusterId to identify the source.

https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateSnapshot
  &ReplicationGroupId=myClusteredRedis
  &SnapshotName=bkup-20150515
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &X-Amz-Credential=<credential>

For more information, see CreateSnapshot in the *Amazon ElastiCache API Reference*.

Related Topics

For more information, see CreateSnapshot in the *Amazon ElastiCache API Reference*.
Creating a Final Backup

You can create a final backup using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Creating a Final Backup (Console)

You can create a final backup when you delete either a Redis cluster (for the API or CLI, a replication group) using the ElastiCache console.

To create a final backup when deleting a Redis cluster, on the delete dialog box (step 5), choose Yes and give the backup a name.

Related Topics

- Using the AWS Management Console (p. 117)
- Deleting a Replication Group (Console) (p. 190)

Creating a Final Backup (AWS CLI)

You can create a final backup when deleting a Redis cluster (for the API or CLI, a replication group) using the AWS CLI.

Topics

- When Deleting a Redis Cluster With No Read Replicas (p. 224)
- When Deleting a Redis Cluster With Read Replicas (p. 224)

When Deleting a Redis Cluster With No Read Replicas

To create a final backup, use the `delete-cache-cluster` AWS CLI operation with the following parameters.

- `--cache-cluster-id` – Name of the cluster being deleted.
- `--final-snapshot-identifier` – Name of the backup.

The following code creates the final backup `bkup-20150515-final` when deleting the cluster `myRedisCluster`.

For Linux, macOS, or Unix:

```bash
aws elasticache delete-cache-cluster \
  --cache-cluster-id myRedisCluster \n  --final-snapshot-identifier bkup-20150515-final
```

For Windows:

```bash
aws elasticache delete-cache-cluster ^
  --cache-cluster-id myRedisCluster ^
  --final-snapshot-identifier bkup-20150515-final
```

For more information, see `delete-cache-cluster` in the AWS CLI Command Reference.

When Deleting a Redis Cluster With Read Replicas

To create a final backup when deleting a replication group, use the `delete-replication-group` AWS CLI operation, with the following parameters:
Creating a Final Backup

- `--replication-group-id` – Name of the replication group being deleted.
- `--final-snapshot-identifier` – Name of the final backup.

The following code takes the final backup `bkup-20150515-final` when deleting the replication group `myReplGroup`.

For Linux, macOS, or Unix:

```bash
aws elasticache delete-replication-group \
  --replication-group-id myReplGroup \
  --final-snapshot-identifier bkup-20150515-final
```

For Windows:

```bash
aws elasticache delete-replication-group ^
  --replication-group-id myReplGroup ^
  --final-snapshot-identifier bkup-20150515-final
```

For more information, see `delete-replication-group` in the AWS CLI Command Reference.

Creating a Final Backup (ElastiCache API)

You can create a final backup when deleting a Redis cluster or replication group using the ElastiCache API.

Topics

- When Deleting a Redis Cluster (p. 225)
- When Deleting a Redis Replication Group (p. 225)

When Deleting a Redis Cluster

To create a final backup, use the `DeleteCacheCluster` ElastiCache API operation with the following parameters.

- `CacheClusterId` – Name of the cluster being deleted.
- `FinalSnapshotIdentifier` – Name of the backup.

The following ElastiCache API operation creates the backup `bkup-20150515-final` when deleting the cluster `myRedisCluster`.

```xml
https://elasticache.us-west-2.amazonaws.com/
  ?Action=DeleteCacheCluster
  &CacheClusterId=myRedisCluster
  &FinalSnapshotIdentifier=bkup-20150515-final
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &X-Amz-Credential=<credential>
```

For more information, see `DeleteCacheCluster` in the Amazon ElastiCache API Reference.

When Deleting a Redis Replication Group

To create a final backup when deleting a replication group, use the `DeleteReplicationGroup` ElastiCache API operation, with the following parameters:
Creating a Final Backup

- **ReplicationGroupId** – Name of the replication group being deleted.
- **FinalSnapshotIdentifier** – Name of the final backup.

The following ElastiCache API operation creates the backup `bkup-20150515-final` when deleting the replication group `myReplGroup`.

```
https://elasticache.us-west-2.amazonaws.com/
    ?Action=DeleteReplicationGroup
    &FinalSnapshotIdentifier=bkup-20150515-final
    &ReplicationGroupId=myReplGroup
    &Version=2015-02-02
    &SignatureVersion=4
    &SignatureMethod=HmacSHA256
    &Timestamp=20150202T192317Z
    &X-Amz-Credential=<credential>
```

For more information, see `DeleteReplicationGroup` in the *Amazon ElastiCache API Reference*. 
Describing Backups

The following procedures show you how to display a list of your backups. If you desire, you can also view the details of a particular backup.

Describing Backups (Console)

To display backups using the AWS Management Console

2. From the navigation pane, choose Backups.
3. Use the Filter list to display manual, automatic, or all backups.
4. To see the details of a particular backup, choose the box to the left of the backup's name.

Describing Backups (AWS CLI)

To display a list of backups and optionally details about a specific backup, use the describe-snapshots CLI operation.

Examples

The following operation uses the parameter --max-records to list up to 20 backups associated with your account. Omitting the parameter --max-records lists up to 50 backups.

```bash
aws elasticache describe-snapshots --max-records 20
```

The following operation uses the parameter --cache-cluster-id to list only the backups associated with the cluster my-cluster.

```bash
aws elasticache describe-snapshots --cache-cluster-id my-cluster
```

The following operation uses the parameter --snapshot-name to display the details of the backup my-backup.

```bash
aws elasticache describe-snapshots --snapshot-name my-backup
```

For more information, see describe-snapshots in the AWS CLI Command Reference.

Describing Backups (ElastiCache API)

To display a list of backups, use the DescribeSnapshots operation.

Examples

The following operation uses the parameter MaxRecords to list up to 20 backups associated with your account. Omitting the parameter MaxRecords lists up to 50 backups.

```xml
https://elasticache.us-west-2.amazonaws.com/
  ?Action=DescribeSnapshots
  &MaxRecords=20
  &SignatureMethod=HmacSHA256
  &SignatureVersion=4
```
The following operation uses the parameter `CacheClusterId` to list all backups associated with the cluster `MyCluster`.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeSnapshots
&CacheClusterId=MyCluster
&SignatureMethod=HmacSHA256
&SignatureVersion=4
&Timestamp=20141201T220302Z
&Version=2014-12-01
&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>
```

The following operation uses the parameter `SnapshotName` to display the details for the backup `MyBackup`.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeSnapshots
&SignatureMethod=HmacSHA256
&SignatureVersion=4
&SnapshotName=MyBackup
&Timestamp=20141201T220302Z
&Version=2014-12-01
&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 `DescribeSnapshots`. 
Copying a Backup

You can make a copy of any backup, whether it was created automatically or manually. You can also export your backup so you can access it from outside ElastiCache. For guidance on exporting your backup, see Exporting a Backup (p. 231).

The following procedures show you how to copy a backup.

Copying a Backup (Console)

To copy a backup (console)

2. To see a list of your backups, from the left navigation pane choose Backups.
3. From the list of backups, choose the box to the left of the name of the backup you want to copy.
4. Choose Copy.
5. In the Create Copy of the Backup? dialog box, do the following:
   a. In the New backup name box, type a name for your new backup.
   b. Leave the optional Target S3 Bucket box blank. This field should only be used to export your backup and requires special S3 permissions. For information on exporting a backup, see Exporting a Backup (p. 231).
   c. Choose Copy.

Copying a Backup (AWS CLI)

To copy a backup, use the copy-snapshot operation.

Parameters

- `--source-snapshot-name` – Name of the backup to be copied.
- `--target-snapshot-name` – Name of the backup's copy.
- `--target-bucket` – Reserved for exporting a backup. Do not use this parameter when making a copy of a backup. For more information, see Exporting a Backup (p. 231).

The following example makes a copy of an automatic backup.

For Linux, macOS, or Unix:

```
aws elasticache copy-snapshot \
    --source-snapshot-name automatic.my-redis-primary-2014-03-27-03-15 \
    --target-snapshot-name my-backup-copy
```

For Windows:

```
aws elasticache copy-snapshot ^
    --source-snapshot-name automatic.my-redis-primary-2014-03-27-03-15 ^
    --target-snapshot-name my-backup-copy
```

For more information, see copy-snapshot in the AWS CLI.
Copying a Backup (ElastiCache API)

To copy a backup, use the CopySnapshot operation with the following parameters:

**Parameters**

- **SourceSnapshotName** – Name of the backup to be copied.
- **TargetSnapshotName** – Name of the backup's copy.
- **TargetBucket** – Reserved for exporting a backup. Do not use this parameter when making a copy of a backup. For more information, see Exporting a Backup (p. 231).

The following example makes a copy of an automatic backup.

**Example**

```text
https://elasticache.us-west-2.amazonaws.com/
?Action=CopySnapshot
&SourceSnapshotName=automatic.my-redis-primary-2014-03-27-03-15
&TargetSnapshotName=my-backup-copy
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&Version=2014-12-01
&X-Amz-Algorithm=AWS4-HMAC-SHA256
&X-Amz-Date=20141201T220302Z
&X-Amz-SignedHeaders=Host
&X-Amz-Signature=<signature>
```

For more information, see CopySnapshot in the Amazon ElastiCache API Reference.
Exporting a Backup

Amazon ElastiCache supports exporting your ElastiCache backup to an Amazon Simple Storage Service (Amazon S3) bucket, which gives you access to it from outside ElastiCache. You can export a backup using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Exporting a backup can be helpful if you need to launch a cluster in another AWS Region. You can export your data in one AWS Region, copy the .rdb file to the new AWS Region, and then use that .rdb file to seed the new cluster instead of waiting for the new cluster to populate through use. For information about seeding a new cluster, see Seeding a New Cluster with an Externally Created Backup (p. 242). Another reason you might want to export your cluster’s data is to use the .rdb file for offline processing.

**Important**

The ElastiCache backup and the Amazon S3 bucket that you want to copy it to must be in the same AWS Region. Though backups copied to an Amazon S3 bucket are encrypted, we strongly recommend that you do not grant others access to the Amazon S3 bucket where you want to store your backups.

Before you can export a backup to an Amazon S3 bucket, you must have an Amazon S3 bucket in the same AWS Region as the backup. Grant ElastiCache access to the bucket. The first two steps show you how to do this.

**Warning**

The following scenarios expose your data in ways that you might not want:

- **When another person has access to the Amazon S3 bucket that you exported your backup to.**

  To control access to your backups, only allow access to the Amazon S3 bucket to those whom you want to access your data. For information about managing access to an Amazon S3 bucket, see Managing Access in the Amazon S3 Developer Guide.

- **When another person has permissions to use the CopySnapshot API operation.**

  Users or groups that have permissions to use the CopySnapshot API operation can create their own Amazon S3 buckets and copy backups to them. To control access to your backups, use an AWS Identity and Access Management (IAM) policy to control who has the ability to use the CopySnapshot API. For more information about using IAM to control the use of ElastiCache API operations, see Identity and Access Management in Amazon ElastiCache (p. 390) in the ElastiCache User Guide.

**Topics**

- Step 1: Create an Amazon S3 Bucket (p. 231)
- Step 2: Grant ElastiCache Access to Your Amazon S3 Bucket (p. 232)
- Step 3: Export an ElastiCache Backup (p. 234)

### Step 1: Create an Amazon S3 Bucket

The following procedure uses the Amazon S3 console to create an Amazon S3 bucket where you export and store your ElastiCache backup.

**To create an Amazon S3 bucket**

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose **Create Bucket**.

3. In **Create a Bucket - Select a Bucket Name and Region**, do the following:
   a. In **Bucket Name**, type a name for your Amazon S3 bucket.
      
      The name of your Amazon S3 bucket must be DNS-compliant. Otherwise, ElastiCache can't access your backup file. The rules for DNS compliance are:
      
      • Names must be at least 3 and no more than 63 characters long.
      • Names must be a series of one or more labels separated by a period (.) where each label:
        • Starts with a lowercase letter or a number.
        • Ends with a lowercase letter or a number.
        • Contains only lowercase letters, numbers, and dashes.
      • Names can't be formatted as an IP address (for example, 192.0.2.0).
   b. From the **Region** list, choose an AWS Region for your Amazon S3 bucket. This AWS Region must be the same AWS Region as the ElastiCache backup you want to export.
   c. Choose **Create**.

For more information about creating an Amazon S3 bucket, see **Creating a Bucket** in the *Amazon Simple Storage Service Console User Guide*.

**Step 2: Grant ElastiCache Access to Your Amazon S3 Bucket**

For ElastiCache to be able to copy a snapshot to an Amazon S3 bucket, grant access to the bucket. You grant ElastiCache access to your Amazon S3 bucket in a different way depending if your bucket is in a default AWS Region or an opt-in AWS Region.

AWS Regions introduced before March 20, 2019, are enabled by default. You can begin working in these AWS Regions immediately. Regions introduced after March 20, 2019, such as Asia Pacific (Hong Kong) and Middle East (Bahrain), are disabled by default. You must enable, or opt in, to these Regions before you can use them, as described in **Managing AWS Regions** in the *AWS General Reference*.

Choose your approach depending on your AWS Region:

- For a default Region, use the procedure in **Grant ElastiCache Access to Your S3 Bucket in a Default Region** (p. 232).
- For an opt-in Region, use the procedure in **Grant ElastiCache Access to Your S3 Bucket in an Opt-In Region** (p. 233).

**Grant ElastiCache Access to Your S3 Bucket in a Default Region**

AWS Regions introduced before March 20, 2019, are enabled by default. You can begin working in these AWS Regions immediately. Regions introduced after March 20, 2019, such as Asia Pacific (Hong Kong) and Middle East (Bahrain), are disabled by default. You must enable, or opt in, to these Regions before you can use them, as described in **Managing AWS Regions** in the *AWS General Reference*.

To create the proper permissions on an Amazon S3 bucket in an AWS Region enabled by default, take the steps described following.

**Warning**

Even though backups copied to an Amazon S3 bucket are encrypted, your data can be accessed by anyone with access to your Amazon S3 bucket. Therefore, we strongly recommend that you set up IAM policies to prevent unauthorized access to this Amazon S3 bucket. For more information, see **Managing Access** in the *Amazon S3 Developer Guide*.

**API Version 2015-02-02**

232
To grant ElastiCache access to an S3 bucket in a default AWS Region

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose the name of the Amazon S3 bucket that you want to copy the backup to. This should be the S3 bucket that you created in Step 1: Create an Amazon S3 Bucket (p. 231).
3. Make sure that the bucket's AWS Region is the same as your ElastiCache backup's AWS Region. If it isn't, return to Step 1: Create an Amazon S3 Bucket (p. 231) and create a new bucket in the same AWS Region as the cluster that you back up.
4. Choose the Permissions tab, choose Access Control List, and under Access for other AWS accounts, choose Add account.
5. In the box, add the AWS Region's canonical ID as shown in the following list:
   - China (Beijing) and China (Ningxia) Regions:
     
     b14d6a125bdf69854ed8ef2e71d8a20b7c490f252229b806e514966e490b8d83
   - AWS GovCloud (US-West) Region:
     
     40fa568277ad703bd160f66ae4f83fc9dfdfd06c2f1b5060aca2442ac3ef8be6
   - All other AWS Regions enabled by default:
     
     540804c33a284a299d2547575ce1010f2312ef3da9b3a053c8bc45bf233e4353

Important
The backup must be exported to an S3 bucket in AWS GovCloud (US).

6. Set the permissions on the bucket by choosing Yes for the following options:
   - List objects
   - Write objects
   - Read bucket permissions
7. Choose Save.

Your Amazon S3 bucket is now ready for you to export an ElastiCache backup to using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Grant ElastiCache Access to Your S3 Bucket in an Opt-In Region

AWS Regions introduced before March 20, 2019, are enabled by default. You can begin working in these AWS Regions immediately. Regions introduced after March 20, 2019, such as Asia Pacific (Hong Kong) and Middle East (Bahrain), are disabled by default. You must enable, or opt in, to these Regions before you can use them, as described in Managing AWS Regions in the AWS General Reference.

To create the proper permissions on an Amazon S3 bucket in an opt-in AWS Region, take the following steps.

To grant ElastiCache access to an S3 bucket in an opt-in AWS Region

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose the name of the Amazon S3 bucket that you want to copy the backup to. This should be the S3 bucket that you created in Step 1: Create an Amazon S3 Bucket (p. 231).
3. Choose the **Permissions** tab. and under **Permissions**, choose **Bucket policy**.
4. Update the policy to grant ElastiCache required permissions to perform operations:
   
   - Add `[ "Service" : "region-full-name.elasticache-snapshot.amazonaws.com" ]` to **Principal**.
   - Add the following permissions required for exporting a snapshot to the Amazon S3 bucket.
     - "s3:PutObject"
     - "s3:GetObject"
     - "s3:ListBucket"
     - "s3:GetBucketAcl"
     - "s3:ListMultipartUploadParts"
     - "s3:ListBucketMultipartUploads"

   The following is an example of what the updated policy might look like.

   ```json
   {
   "Version": "2012-10-17",
   "Id": "Policy15397346",
   "Statement": [
   {
   "Sid": "Stmt15399483",
   "Effect": "Allow",
   "Principal": {
   "Service": "aws-opt-in-region.elasticache-snapshot.amazonaws.com"
   },
   "Action": [
   "s3:PutObject",
   "s3:GetObject",
   "s3:ListBucket",
   "s3:GetBucketAcl",
   "s3:ListMultipartUploadParts",
   "s3:ListBucketMultipartUploads"
   ],
   "Resource": [
   "arn:aws:s3:::example-bucket",
   "arn:aws:s3:::example-bucket/*"
   ]
   }
   ]
   }
   
   Step 3: Export an ElastiCache Backup

   Now you've created your S3 bucket and granted ElastiCache permissions to access it. Next, you can use the ElastiCache console, the AWS CLI, or the ElastiCache API to export your snapshot to it. The following assumes that you have the following additional S3 specific IAM permissions.

   ```json
   {
   "Statement": [
   {
   "Effect": "Allow",
   "Action": [
   "s3:GetBucketLocation",
   "s3:ListAllMyBuckets",
   "s3:PutObject",
   "s3:GetObject",
   "s3:DeleteObject"
   ],
   "Resource": [
   "arn:aws:s3:::example-bucket",
   "arn:aws:s3:::example-bucket/*"
   ]
   }
   ]
   }
   ```
Exporting an ElastiCache Backup (Console)

The following process uses the ElastiCache console to export a backup to an Amazon S3 bucket so that you can access it from outside ElastiCache. The Amazon S3 bucket must be in the same AWS Region as the ElastiCache backup.

To export an ElastiCache backup to an Amazon S3 bucket

2. To see a list of your backups, from the left navigation pane choose Backups.
3. From the list of backups, choose the box to the left of the name of the backup you want to export.
4. Choose Copy.
5. In Create a Copy of the Backup?, do the following:
   a. In New backup name box, type a name for your new backup.
      The name must be between 1 and 1,000 characters and able to be UTF-8 encoded.
      ElastiCache adds an instance identifier and .rdb to the value that you enter here. For example, if you enter my-exported-backup, ElastiCache creates my-exported-backup-0001.rdb.
   b. From the Target S3 Location list, choose the name of the Amazon S3 bucket that you want to copy your backup to (the bucket that you created in Step 1: Create an Amazon S3 Bucket (p. 231)).
      The Target S3 Location must be an Amazon S3 bucket in the backup's AWS Region with the following permissions for the export process to succeed.
      • Object access – Read and Write.
      • Permissions access – Read.
      For more information, see Step 2: Grant ElastiCache Access to Your Amazon S3 Bucket (p. 232).
   c. Choose Copy.

Note
If your S3 bucket does not have the permissions needed for ElastiCache to export a backup to it, you receive one of the following error messages. Return to Step 2: Grant ElastiCache Access to Your Amazon S3 Bucket (p. 232) to add the permissions specified and retry exporting your backup.

• ElastiCache has not been granted READ permissions %s on the S3 Bucket.
  Solution: Add Read permissions on the bucket.
• ElastiCache has not been granted WRITE permissions %s on the S3 Bucket.
  Solution: Add Write permissions on the bucket.
• ElastiCache has not been granted READ_ACP permissions %s on the S3 Bucket.
Solution: Add Read for Permissions access on the bucket.

If you want to copy your backup to another AWS Region, use Amazon S3 to copy it. For more information, see Copying an Object in the Amazon Simple Storage Service Console User Guide.

Exporting an ElastiCache Backup (AWS CLI)

Export the backup to an Amazon S3 bucket using the `copy-snapshot` CLI operation with the following parameters:

**Parameters**

- `--source-snapshot-name` – Name of the backup to be copied.
- `--target-snapshot-name` – Name of the backup's copy.

The name must be between 1 and 1,000 characters and able to be UTF-8 encoded.

ElastiCache adds an instance identifier and `.rdb` to the value you enter here. For example, if you enter `my-exported-backup`, ElastiCache creates `my-exported-backup-0001.rdb`.

- `--target-bucket` – Name of the Amazon S3 bucket where you want to export the backup. A copy of the backup is made in the specified bucket.

The `--target-bucket` must be an Amazon S3 bucket in the backup's AWS Region with the following permissions for the export process to succeed.

- Object access – Read and Write.
- Permissions access – Read.

For more information, see Step 2: Grant ElastiCache Access to Your Amazon S3 Bucket (p. 232).

The following operation copies a backup to `my-s3-bucket`.

For Linux, macOS, or Unix:

```
aws elasticache copy-snapshot \
  --source-snapshot-name automatic.my-redis-primary-2016-06-27-03-15 \
  --target-snapshot-name my-exported-backup \
  --target-bucket my-s3-bucket
```

For Windows:

```
aws elasticache copy-snapshot ^
  --source-snapshot-name automatic.my-redis-primary-2016-06-27-03-15 ^
  --target-snapshot-name my-exported-backup ^
  --target-bucket my-s3-bucket
```

Note

If your S3 bucket does not have the permissions needed for ElastiCache to export a backup to it, you receive one of the following error messages. Return to Step 2: Grant ElastiCache Access to Your Amazon S3 Bucket (p. 232) to add the permissions specified and retry exporting your backup.

- ElastiCache has not been granted READ permissions %s on the S3 Bucket.

  Solution: Add Read permissions on the bucket.

- ElastiCache has not been granted WRITE permissions %s on the S3 Bucket.
Solution: Add Write permissions on the bucket.

- ElastiCache has not been granted READ_ACP permissions on the S3 Bucket.

Solution: Add Read for Permissions access on the bucket.

For more information, see copy-snapshot in the AWS CLI Command Reference.

If you want to copy your backup to another AWS Region, use Amazon S3 copy. For more information, see Copying an Object in the Amazon Simple Storage Service Console User Guide.

Exporting an ElastiCache Backup (ElastiCache API)

Export the backup to an Amazon S3 bucket using the CopySnapshot API operation with these parameters.

Parameters

- SourceSnapshotName – Name of the backup to be copied.
- TargetSnapshotName – Name of the backup's copy.

The name must be between 1 and 1,000 characters and able to be UTF-8 encoded.

ElastiCache adds an instance identifier and .rdb to the value that you enter here. For example, if you enter my-exported-backup, you get my-exported-backup-0001.rdb.

- TargetBucket – Name of the Amazon S3 bucket where you want to export the backup. A copy of the backup is made in the specified bucket.

The TargetBucket must be an Amazon S3 bucket in the backup's AWS Region with the following permissions for the export process to succeed.

- Object access – Read and Write.
- Permissions access – Read.

For more information, see Step 2: Grant ElastiCache Access to Your Amazon S3 Bucket (p. 232).

The following example makes a copy of an automatic backup to the Amazon S3 bucket my-s3-bucket.

Example

```
https://elasticache.us-west-2.amazonaws.com/
?Action=CopySnapshot
&SourceSnapshotName=automatic.my-redis-primary-2016-06-27-03-15
&TargetBucket=my-s3-bucket
&TargetSnapshotName=my-backup-copy
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&Version=2016-01-01
&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>
```

Note

If your S3 bucket does not have the permissions needed for ElastiCache to export a backup to it, you receive one of the following error messages. Return to Step 2: Grant ElastiCache Access.
to Your Amazon S3 Bucket (p. 232) to add the permissions specified and retry exporting your backup.

- ElastiCache has not been granted READ permissions on the S3 Bucket.
  
  **Solution:** Add Read permissions on the bucket.

- ElastiCache has not been granted WRITE permissions on the S3 Bucket.
  
  **Solution:** Add Write permissions on the bucket.

- ElastiCache has not been granted READ_ACP permissions on the S3 Bucket.
  
  **Solution:** Add Read for Permissions access on the bucket.

For more information, see CopySnapshot in the Amazon ElastiCache API Reference.

If you want to copy your backup to another AWS Region, use Amazon S3 copy to copy the exported backup to the Amazon S3 bucket in another AWS Region. For more information, see Copying an Object in the Amazon Simple Storage Service Console User Guide.
Restoring From a Backup with Optional Cluster Resizing

You can restore the data from a Redis .rdb backup file to a new cluster at any time.

The Amazon ElastiCache for Redis restore process supports the following:

• Upgrading from a Redis (cluster mode disabled) cluster to a Redis (cluster mode enabled) cluster running Redis version 3.2.4.
• Migrating from one or more .rdb backup files you created from your self-managed Redis clusters to a single ElastiCache for Redis (cluster mode enabled) cluster.

The .rdb files must be put in S3 to perform the restore.
• Specifying a number of shards (API/CLI: node groups) in the new cluster that is different from the number of shards in the cluster that was used to create the backup file.
• Specifying a different node type for the new cluster—larger or smaller. If scaling to a smaller node type, be sure that the new node type has sufficient memory for your data and Redis overhead. For more information, see Choosing Your Node Size (p. 76).
• Configuring the slots of the new Redis (cluster mode enabled) cluster differently than in the cluster that was used to create the backup file.

Important

• You cannot restore from a backup created using a Redis (cluster mode enabled) cluster to a Redis (cluster mode disabled) cluster.
• Redis (cluster mode enabled) clusters do not support multiple databases. Therefore, when restoring to a Redis (cluster mode enabled) your restore fails if the .rdb file references more than one database.

Whether you make any changes when restoring a cluster from a backup is governed by choices that you make. You make these choices in the Restore Cluster dialog box when using the ElastiCache console to restore. You make these choices by setting parameter values when using the AWS CLI or ElastiCache API to restore.

During the restore operation, ElastiCache creates the new cluster, and then populates it with data from the backup file. When this process is complete, the Redis cluster is warmed up and ready to accept requests.

Important
Before you proceed, be sure you have created a backup of the cluster you want to restore from. For more information, see Making Manual Backups (p. 218).
If you want to restore from an externally created backup, see Seeding a New Cluster with an Externally Created Backup (p. 242).

The following procedures show you how to restore a backup to a new cluster using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Topics
• Restoring From a Backup (Console) (p. 240)
• Restoring From a Backup (AWS CLI) (p. 241)
• Restoring From a Backup (ElastiCache API) (p. 241)
Restoring From a Backup (Console)

You can restore a Redis backup in two ways. You can restore to a single-node Redis (cluster mode disabled) cluster. Or you can restore to a Redis cluster with read replicas (a replication group), either Redis (cluster mode disabled) or Redis (cluster mode enabled).

To restore a backup to a new cluster (console)

2. From the navigation pane, choose Backups.
3. In the list of backups, choose the box to the left of the backup name you want to restore from.
4. Choose Restore.
5. Complete the Restore Cluster dialog box. Be sure to complete all the "Required" fields and any of the others you want to change from the defaults.

Redis (Cluster Mode Disabled)

1. Cluster ID – Required. The name of the new cluster.
2. Engine version compatibility – The ElastiCache for Redis engine version you want to run.
3. Cluster mode enabled (scale out) – Choose this to convert your Redis (cluster mode disabled) cluster to a Redis (cluster mode enabled). The engine version becomes 3.2.4.

If you choose Cluster mode enabled (scale out):
   a. Choose the number of shards you want in the new cluster (API/CLI: node groups).
   b. Choose the number of read replicas you want in each shard.
   c. Distribute your keys among the slots as you desire.
4. Node Type – Specify the node type you want for the new cluster.
5. Availability zone(s) – Specify how you want the cluster's Availability Zones selected.
6. Port – Change this only if you want the new cluster to use a different port.
7. Choose a VPC – Choose the VPC in which to create this cluster.
8. Parameter Group – Choose a parameter group that reserves sufficient memory for Redis overhead for the node type you selected.

Redis (Cluster Mode Enabled)

1. Cluster ID – Required. The name of the new cluster.
2. Cluster mode enabled (scale out) – Choose this for a Redis (cluster mode enabled) cluster. Clear it for a Redis (cluster mode disabled) cluster.
3. Node Type – Specify the node type you want for the new cluster.
4. Number of Shards – Choose the number of shards you want in the new cluster (API/CLI: node groups).
5. Replicas per Shard – Choose the number of read replica nodes you want in each shard.
6. Slots and keyspaces – Choose how you want keys distributed among the shards. If you choose to specify the key distributions complete the table specifying the key ranges for each shard.
7. Availability zone(s) – Specify how you want the cluster's Availability Zones selected.
8. Port – Change this only if you want the new cluster to use a different port.
9. Choose a VPC – Choose the VPC in which to create this cluster.
10 Parameter Group – Choose a parameter group that reserves sufficient memory for Redis overhead for the node type you selected.

6. When the settings are as you want them, choose Launch Cluster.

Restoring From a Backup (AWS CLI)

You can restore a Redis (cluster mode disabled) backup in two ways. You can restore to a single-node Redis (cluster mode disabled) cluster using the AWS CLI operation create-cache-cluster. Or you can restore to a Redis cluster with read replicas (a replication group). To do the latter, you can use either Redis (cluster mode disabled) or Redis (cluster mode enabled) with the AWS CLI operation create-replication-group. In this case, you seed the restore with a Redis .rdb file.

When using either the create-cache-cluster or create-replication-group operation, be sure to include the parameter --snapshot-name or --snapshot-arns to seed the new cluster or replication group with the data from the backup.

For more information, see the following:

- Creating a Cluster (AWS CLI) (p. 87) in the ElastiCache User Guide.
- create-cache-cluster in the AWS CLI Command Reference.

- Creating a Redis Replication Group from Scratch (p. 168) in the ElastiCache User Guide.
- create-replication-group in the AWS CLI Command Reference.

Restoring From a Backup (ElastiCache API)

You can restore a Redis backup to either a single-node Redis (cluster mode disabled) cluster using the ElastiCache API operation CreateCacheCluster or to a Redis cluster with read replicas (replication group)— either Redis (cluster mode disabled) or Redis (cluster mode enabled) using the ElastiCache API operation CreateReplicationGroup and seeding it with a Redis .rdb file.

When using either the CreateCacheCluster or CreateReplicationGroup operation, be sure to include the parameter SnapshotName or SnapshotArns to seed the new cluster or replication group with the data from the backup.

For more information, see the following:

- CreateCacheCluster in the ElastiCache API Reference.

- Creating a Redis Replication Group from Scratch (p. 168) in the ElastiCache User Guide.
- CreateReplicationGroup in the ElastiCache API Reference.
Seeding a New Cluster with an Externally Created Backup

When you create a new Redis cluster, you can seed it with data from a Redis .rdb backup file. Seeding the cluster is useful if you currently manage a Redis instance outside of ElastiCache and want to populate your new ElastiCache for Redis cluster with your existing Redis data.

To seed a new Redis cluster from a Redis backup created within Amazon ElastiCache, see Restoring From a Backup with Optional Cluster Resizing (p. 239).

When you use a Redis .rdb file to seed a new Redis cluster, you can do the following:

- Upgrade from a nonpartitioned cluster to a Redis (cluster mode enabled) cluster running Redis version 3.2.4.
- Specify a number of shards (called node groups in the API and CLI) in the new cluster. This number can be different from the number of shards in the cluster that was used to create the backup file.
- Specify a different node type for the new cluster—larger or smaller than that used in the cluster that made the backup. If you scale to a smaller node type, be sure that the new node type has sufficient memory for your data and Redis overhead. For more information, see Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464).
- Distribute your keys in the slots of the new Redis (cluster mode enabled) cluster differently than in the cluster that was used to create the backup file.

**Note**
You can't seed a Redis (cluster mode disabled) cluster from an .rdb file created from a Redis (cluster mode enabled) cluster.

**Important**
- You must ensure that your Redis backup data doesn't exceed the resources of the node. For example, you can't upload an .rdb file with 5 GB of Redis data to a cache.m3.medium node that has 2.9 GB of memory.

If the backup is too large, the resulting cluster has a status of `restore-failed`. If this happens, you must delete the cluster and start over.

For a complete listing of node types and specifications, see Redis Node-Type Specific Parameters (p. 336) and Amazon ElastiCache Product Features and Details.
- You can encrypt a Redis .rdb file with Amazon S3 server-side encryption (SSE-S3) only. For more information, see Protecting Data Using Server-Side Encryption.

Following, you can find topics that walk you through migrating your Redis cluster from outside ElastiCache for Redis to ElastiCache for Redis.

**Migrating to ElastiCache for Redis**
- Step 1: Create a Redis Backup (p. 243)
- Step 2: Create an Amazon S3 Bucket and Folder (p. 243)
- Step 3: Upload Your Backup to Amazon S3 (p. 244)
- Step 4: Grant ElastiCache Read Access to the .rdb File (p. 244)
- Step 5: Seed the ElastiCache Cluster With the .rdb File Data (p. 247)
Step 1: Create a Redis Backup

To create the Redis backup to seed your ElastiCache for Redis instance

1. Connect to your existing Redis instance.
2. Run either the Redis `BGSAVE` or `SAVE` operation to create a backup. Note where your .rdb file is located.

   `BGSAVE` is asynchronous and does not block other clients while processing. For more information, see `BGSAVE` at the Redis website.

   `SAVE` is synchronous and blocks other processes until finished. For more information, see `SAVE` at the Redis website.

For additional information on creating a backup, see Redis Persistence at the Redis website.

Step 2: Create an Amazon S3 Bucket and Folder

When you have created the backup file, you need to upload it to a folder within an Amazon S3 bucket. To do that, you must first have an Amazon S3 bucket and folder within that bucket. If you already have an Amazon S3 bucket and folder with the appropriate permissions, you can skip to Step 3: Upload Your Backup to Amazon S3 (p. 244).

To create an Amazon S3 bucket

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Follow the instructions for creating an Amazon S3 bucket in Creating a Bucket in the Amazon Simple Storage Service Console User Guide.

   The name of your Amazon S3 bucket must be DNS-compliant. Otherwise, ElastiCache can't access your backup file. The rules for DNS compliance are:
   - Names must be at least 3 and no more than 63 characters long.
   - Names must be a series of one or more labels separated by a period (.) where each label:
     - Starts with a lowercase letter or a number.
     - Ends with a lowercase letter or a number.
     - Contains only lowercase letters, numbers, and dashes.
   - Names can't be formatted as an IP address (for example, 192.0.2.0).

   We strongly recommend that you create your Amazon S3 bucket in the same AWS Region as your new ElastiCache for Redis cluster. This approach makes sure that the highest data transfer speed when ElastiCache reads your .rdb file from Amazon S3.

   **Note**
   To keep your data as secure as possible, make the permissions on your Amazon S3 bucket as restrictive as you can. At the same time, the permissions still need to allow the bucket and its contents to be used to seed your new Redis cluster.

To add a folder to an Amazon S3 bucket

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose the name of the bucket to upload your .rdb file to.
3. Choose Create folder.
4. Enter a name for your new folder.
5. Choose Save.

Make note of both the bucket name and the folder name.

### Step 3: Upload Your Backup to Amazon S3

Now, upload the .rdb file that you created in Step 1: Create a Redis Backup (p. 243). You upload it to the Amazon S3 bucket and folder that you created in Step 2: Create an Amazon S3 Bucket and Folder (p. 243). For more information on this task, see Add an Object to a Bucket. Between steps 2 and 3, choose the name of the folder you created.

#### To upload your .rdb file to an Amazon S3 folder

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose the name of the Amazon S3 bucket you created in Step 2.
3. Choose the name of the folder you created in Step 2.
5. Choose Add files.
6. Browse to find the file or files you want to upload, then choose the file or files. To choose multiple files, hold down the Ctrl key while choosing each file name.
7. Choose Open.
8. Confirm the correct file or files are listed in the Upload dialog box, and then choose Upload.

Note the path to your .rdb file. For example, if your bucket name is myBucket and the path is myFolder/redis.rdb, enter myBucket/myFolder/redis.rdb. You need this path to seed the new cluster with the data in this backup.

For additional information, see Bucket Restrictions and Limitations in the Amazon Simple Storage Service Developer Guide.

### Step 4: Grant ElastiCache Read Access to the .rdb File

Now, grant ElastiCache read access to your .rdb backup file. You grant ElastiCache access to your backup file in a different way depending if your bucket is in a default AWS Region or an opt-in AWS Region.

AWS Regions introduced before March 20, 2019, are enabled by default. You can begin working in these AWS Regions immediately. Regions introduced after March 20, 2019, such as Asia Pacific (Hong Kong) and Middle East (Bahrain), are disabled by default. You must enable, or opt in, to these Regions before you can use them, as described in Managing AWS Regions in the AWS General Reference.

Choose your approach depending on your AWS Region:

- For a default Region, use the procedure in Grant ElastiCache Read Access to the .rdb File in a Default Region (p. 244).
- For an opt-in Region, use the procedure in Grant ElastiCache Read Access to the .rdb File in an Opt-In Region (p. 245).

#### Grant ElastiCache Read Access to the .rdb File in a Default Region

AWS Regions introduced before March 20, 2019, are enabled by default. You can begin working in these AWS Regions immediately. Regions introduced after March 20, 2019, such as Asia Pacific (Hong Kong)
and Middle East (Bahrain), are disabled by default. You must enable, or opt in, to these Regions before you can use them, as described in Managing AWS Regions in the AWS General Reference.

To grant ElastiCache read access to the backup file in an AWS Region enabled by default

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose the name of the S3 bucket that contains your .rdb file.
3. Choose the name of the folder that contains your .rdb file.
4. Choose the name of your .rdb backup file. The name of the selected file appears above the tabs at the top of the page.

5. Choose Permissions.

6. If aws-scs-s3-readonly or one of the canonical IDs in the following list is not listed as a user, do the following:
   a. Under Access for other AWS accounts, choose Add account.
   b. In the box, add the AWS Region's canonical ID as shown in the following list:
      • China (Beijing) and China (Ningxia) Regions:
        b14d6a125bdf69854ed8ef2e71d8a20b7c490f252229b806e514966e490b8d83
      • AWS GovCloud (US-West) Region:
        40fa568277ad703bd160f66ae4f83fc9dfdfd06c2f1b5060ca22442ac3ef8be6

      **Important**
      The backup must be located in an S3 bucket in AWS GovCloud (US) for you to download it to a Redis cluster in AWS GovCloud (US).

      • All other AWS Regions enabled by default:
        540804c33a284a299d2547575ce1010f2312ef3da9b3a053c8bc45bf233e4353
   c. Set the permissions on the bucket by choosing Yes for the following:
      • Read object
      • Read object permissions
   d. Choose Save.
7. Choose Overview, and then choose Download.

Grant ElastiCache Read Access to the .rdb File in an Opt-In Region

AWS Regions introduced before March 20, 2019, are enabled by default. You can begin working in these AWS Regions immediately. Regions introduced after March 20, 2019, such as Asia Pacific (Hong Kong)
and Middle East (Bahrain), are disabled by default. You must enable, or opt in, to these Regions before you can use them, as described in Managing AWS Regions in the AWS General Reference.

To grant ElastiCache read access to the backup file in an opt-in AWS Region

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose the name of the S3 bucket that contains your .rdb file.
3. Choose the name of the folder that contains your .rdb file.
4. Choose the name of your .rdb backup file. The name of the selected file appears above the tabs at the top of the page.
5. Choose the Permissions tab.
6. Under Permissions, choose Bucket policy.
7. Update the policy to grant ElastiCache required permissions to perform operations:
   • Add [ "Service" : "region-full-name.elasticache-snapshot.amazonaws.com" ] to Principal.
   • Add the following permissions required for exporting a snapshot to the Amazon S3 bucket:
     • "s3:GetObject"
     • "s3:ListBucket"
     • "s3:GetBucketAcl"

The following is an example of what the updated policy might look like.

```json
{
  "Version": "2012-10-17",
  "Id": "Policy15397346",
  "Statement": [
    {
      "Sid": "Stmt15399483",
      "Effect": "Allow",
      "Principal": {
        "Service": "ap-east-1.elasticache-snapshot.amazonaws.com"
      },
      "Action": [
        "s3:GetObject",
        "s3:ListBucket",
        "s3:GetBucketAcl"
      ],
      "Resource": [
        "arn:aws:s3:::example-bucket",
        "arn:aws:s3:::example-bucket/backup1.rdb",
        "arn:aws:s3:::example-bucket/backup2.rdb"
      ]
    }
  ]
}
```
Step 5: Seed the ElastiCache Cluster With the .rdb File Data

Now you are ready to create an ElastiCache cluster and seed it with the data from the .rdb file. To create the cluster, follow the directions at Creating a Cluster (p. 74) or Creating a Redis Replication Group from Scratch (p. 168). Be sure to choose Redis as your cluster engine.

The method you use to tell ElastiCache where to find the Redis backup you uploaded to Amazon S3 depends on the method you use to create the cluster:

Seed the ElastiCache for Redis cluster or replication group with the .rdb file data

- **Using the ElastiCache console**

  After you choose the Redis engine, expand the Advanced Redis settings section and locate Import data to cluster. In the Seed RDB file S3 location box, type in the Amazon S3 path for the files(s). If you have multiple .rdb files, type in the path for each file in a comma separated list. The Amazon S3 path looks something like myBucket/myFolder/myBackupFilename.rdb.

- **Using the AWS CLI**

  If you use the create-cache-cluster or the create-replication-group operation, use the parameter --snapshot-arns to specify a fully qualified ARN for each .rdb file. For example, arn:aws:s3:::myBucket/myFolder/myBackupFilename.rdb. The ARN must resolve to the backup files you stored in Amazon S3.

- **Using the ElastiCache API**

  If you use the CreateCacheCluster or the CreateReplicationGroup ElastiCache API operation, use the parameter SnapshotArns to specify a fully qualified ARN for each .rdb file. For example, arn:aws:s3:::myBucket/myFolder/myBackupFilename.rdb. The ARN must resolve to the backup files you stored in Amazon S3.

**Important**

When seeding a Redis (cluster mode enabled) cluster, you must configure each node group (shard) in the new cluster or replication group. Use the parameter --node-group-configuration (API: NodeGroupConfiguration) to do this. For more information, see the following:

- CLI: create-replication-group in the AWS CLI Reference
- API: CreateReplicationGroup in the ElastiCache API Reference

During the process of creating your cluster, the data in your Redis backup is written to the cluster. You can monitor the progress by viewing the ElastiCache event messages. To do this, see the ElastiCache console and choose Cache Events. You can also use the AWS ElastiCache command line interface or ElastiCache API to obtain event messages. For more information, see Viewing ElastiCache Events (p. 426).
Tagging Backups

Cost allocation tags are a means of tracking your costs across multiple AWS services by grouping your expenses on invoices by tag values. To learn more about cost allocation tags, see `Use Cost Allocation Tags`.

Using the ElastiCache console, the AWS CLI, or ElastiCache API you can add, list, modify, remove, or copy cost allocation tags on your backups. For more information, see `Monitoring Costs with Cost Allocation Tags (p. 432)`.
Deleting a Backup

An automatic backup is automatically deleted when its retention limit expires. If you delete a cluster, all of its automatic backups are also deleted. If you delete a replication group, all of the automatic backups from the clusters in that group are also deleted.

ElastiCache provides a deletion API that lets you delete a backup at any time, regardless of whether the backup was created automatically or manually. (Since manual backups do not have a retention limit, manual deletion is the only way to remove them.)

You can delete a backup using the ElastiCache console, the AWS CLI, or the ElastiCache API.

Deleting a Backup (Console)

The following procedure deletes a backup using the ElastiCache console.

To delete a backup

2. In the navigation pane, choose Backups.
   The Backups screen appears with a list of your backups.
3. Choose the box to the left of the name of the backup you want to delete.
4. Choose Delete.
5. If you want to delete this backup, choose Delete on the Delete Backup confirmation screen. The status changes to deleting.

Deleting a Backup (AWS CLI)

Use the delete-snapshot AWS CLI operation with the following parameter to delete a backup.

- --snapshot-name – Name of the backup to be deleted.

The following code deletes the backup myBackup.

```bash
aws elasticache delete-snapshot --snapshot-name myBackup
```

For more information, see delete-snapshot in the AWS CLI Command Reference.

Deleting a Backup (ElastiCache API)

Use the DeleteSnapshot API operation with the following parameter to delete a backup.

- SnapshotName – Name of the backup to be deleted.

The following code deletes the backup myBackup.

```xml
https://elasticache.us-west-2.amazonaws.com/?Action=DeleteSnapshot&SignatureVersion=4
```

API Version 2015-02-02
## Append Only Files (AOF) in ElastiCache for Redis

By default, the data in a Redis node on ElastiCache resides only in memory and isn't persistent. If a node is rebooted, or if the underlying physical server experiences a hardware failure, the data in the cache is lost.

If you require data durability, you can enable the Redis append-only file feature (AOF). When this feature is enabled, the node writes all of the commands that change cache data to an append-only file. When a node is rebooted and the cache engine starts, the AOF is "replayed." The result is a warm Redis cache with all of the data intact.

AOF is disabled by default. To enable AOF for a cluster running Redis, you must create a parameter group with the `appendonly` parameter set to yes. You then assign that parameter group to your cluster. You can also modify the `appendfsync` parameter to control how often Redis writes to the AOF file.

### Important

Append-only files (AOF) aren't supported for cache.t1.micro and cache.t2.* nodes. For nodes of these types, the `appendonly` parameter value is ignored.

For Multi-AZ replication groups, AOF isn't enabled.

AOF isn't supported on Redis versions 2.8.22 and later.

### Warning

AOF can't protect against all failure scenarios. For example, if a node fails due to a hardware fault in an underlying physical server, ElastiCache provisions a new node on a different server. In this case, the AOF file is no longer be available and can't be used to recover the data. Thus, Redis restarts with a cold cache.

For greater reliability and faster recovery, we recommend that you create one or more read replicas in different Availability Zones for your cluster. Enable Multi-AZ on your replication group instead of using AOF. AOF isn't enabled for Multi-AZ replication groups.

For more information on mitigating failures, see [Mitigating Failures when Running Redis](p. 454).

For more information, see the following:

- Redis Specific Parameters (p. 315)
- Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148)
- Mitigating Failures (p. 454)

## Scaling ElastiCache for Redis 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 Redis Clusters

<table>
<thead>
<tr>
<th>Action</th>
<th>Redis (cluster mode disabled)</th>
<th>Redis (cluster mode enabled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling in</td>
<td>Removing Nodes from a Cluster (p. 110)</td>
<td>Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)</td>
</tr>
<tr>
<td>Changing node types</td>
<td><strong>To a larger node type</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scaling Up Single-Node Clusters for Redis (Cluster Mode Disabled) (p. 254)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scaling Up Redis Clusters with Replicas (p. 265)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>To a smaller node type</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scaling Down Single-Node Redis Clusters (p. 259)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Scaling Down Redis Clusters with Replicas (p. 271)</td>
<td></td>
</tr>
<tr>
<td>Changing the number of node groups</td>
<td>Not supported for Redis (cluster mode disabled) clusters</td>
<td>Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)</td>
</tr>
</tbody>
</table>

**Topics**
- Scaling Clusters for Redis (Cluster Mode Disabled) (p. 252)
- Scaling Clusters in Redis (Cluster Mode Enabled) (p. 279)
Scaling Clusters for Redis (Cluster Mode Disabled)

Redis (cluster mode disabled) clusters can be a single-node cluster with 0 shards or multi-node clusters with 1 shard. Single-node clusters use the one node for both reads and writes. Multi-node clusters always have 1 node as the read/write primary node with 0 to 5 read-only replica nodes.

Contents

• Scaling Single-Node Clusters for Redis (Cluster Mode Disabled) (p. 252)
  • Scaling Up Single-Node Clusters for Redis (Cluster Mode Disabled) (p. 254)
    • Scaling Up Single-Node Clusters for Redis (Cluster Mode Disabled) (Console) (p. 255)
    • Scaling Up Single-Node Redis Cache Clusters (AWS CLI) (p. 255)
    • Scaling Up Single-Node Redis Cache Clusters (ElastiCache API) (p. 257)
  • Scaling Down Single-Node Redis Clusters (p. 259)
    • Scaling Down a Single-Node Redis Cluster (Console) (p. 259)
    • Scaling Down Single-Node Redis Cache Clusters (AWS CLI) (p. 260)
    • Scaling Down Single-Node Redis Cache Clusters (ElastiCache API) (p. 262)
• Scaling Redis (Cluster Mode Disabled) Clusters with Replica Nodes (p. 264)
  • Scaling Up Redis Clusters with Replicas (p. 265)
    • Scaling Up a Redis Cluster with Replicas (Console) (p. 266)
    • Scaling Up a Redis Replication Group (AWS CLI) (p. 266)
    • Scaling Up a Redis Replication Group (ElastiCache API) (p. 269)
  • Scaling Down Redis Clusters with Replicas (p. 271)
    • Scaling Down a Redis Cluster with Replicas (Console) (p. 271)
    • Scaling Down a Redis Replication Group (AWS CLI) (p. 266)
    • Scaling Down a Redis Replication Group (ElastiCache API) (p. 274)
• Increasing Read Capacity (p. 277)
• Decreasing Read Capacity (p. 278)

Scaling Single-Node Clusters for Redis (Cluster Mode Disabled)

Redis (cluster mode disabled) nodes must be large enough to contain all the cache's data plus Redis overhead. To change the data capacity of your Redis (cluster mode disabled) cluster, you must scale vertically; scaling up to a larger node type to increase data capacity, or scaling down to a smaller node type to reduce data capacity.

The ElastiCache for Redis scaling up process is designed to make a best effort to retain your existing data and requires successful Redis replication. For Redis (cluster mode disabled) Redis clusters, we recommend that sufficient memory be made available to Redis.

You cannot partition your data across multiple Redis (cluster mode disabled) clusters. However, if you only need to increase or decrease your cluster's read capacity, you can create a Redis (cluster mode disabled) cluster with replica nodes and add or remove read replicas. To create a Redis (cluster mode disabled) cluster with replica nodes using your single-node Redis cache cluster as the primary cluster, see Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79).

After you create the cluster with replicas, you can increase read capacity by adding read replicas. Later, if you need to, you can reduce read capacity by removing read replicas. For more information, see Increasing Read Capacity (p. 277) or Decreasing Read Capacity (p. 278).
In addition to being able to scale read capacity, Redis (cluster mode disabled) clusters with replicas provide other business advantages. For more information, see High Availability Using Replication Groups (p. 142).

**Important**
If your parameter group uses reserved-memory to set aside memory for Redis overhead, before you begin scaling be sure that you have a custom parameter group that reserves the correct amount of memory for your new node type. Alternatively, you can modify a custom parameter group so that it uses reserved-memory-percent and use that parameter group for your new cluster.
If you’re using reserved-memory-percent, doing this is not necessary. For more information, see Managing Reserved Memory (p. 466).

**Topics**
- Scaling Up Single-Node Clusters for Redis (Cluster Mode Disabled) (p. 254)
- Scaling Down Single-Node Redis Clusters (p. 259)
Scaling Up Single-Node Clusters for Redis (Cluster Mode Disabled)

When you scale up a single-node Redis cluster, ElastiCache performs the following process, whether you use the ElastiCache console, the AWS CLI, or the ElastiCache API.

1. A new cache cluster with the new node type is spun up in the same Availability Zone as the existing cache cluster.
2. The cache data in the existing cache cluster is copied to the new cache cluster. How long this process takes depends upon your node type and how much data is in the cache cluster.
3. Reads and writes are now served using the new cache cluster. Because the new cache cluster's endpoints are the same as they were for the old cache cluster, you do not need to update the endpoints in your application. You will notice a brief interruption of reads and writes from the primary node while the DNS entry is updated.
4. ElastiCache deletes the old cache cluster.

As shown in the following table, your Redis scale-up operation is blocked if you have an engine upgrade scheduled for the next maintenance window. For more information on Maintenance Windows, see Managing Maintenance (p. 59).

**Blocked Redis operations**

<table>
<thead>
<tr>
<th>Pending Operations</th>
<th>Blocked Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale up</td>
<td>Immediate engine upgrade</td>
</tr>
<tr>
<td>Engine upgrade</td>
<td>Immediate scale up</td>
</tr>
<tr>
<td>Scale up and engine upgrade</td>
<td>Immediate scale up</td>
</tr>
<tr>
<td></td>
<td>Immediate engine upgrade</td>
</tr>
</tbody>
</table>

If you have a pending operation that is blocking you, you can do one of the following.

- Schedule your Redis scale-up operation for the next maintenance window by clearing the **Apply immediately** check box (CLI use: `--no-apply-immediately`, API use: `ApplyImmediately=false`).
- Wait until your next maintenance window (or after) to perform your Redis scale up operation.
- Add the Redis engine upgrade to this cache cluster modification with the **Apply Immediately** check box chosen (CLI use: `--apply-immediately`, API use: `ApplyImmediately=true`). This unblocks your scale up operation by causing the engine upgrade to be performed immediately.

You can scale up a single-node Redis (cluster mode disabled) cluster using the ElastiCache console, the AWS CLI, or ElastiCache API.

**Important**

If your parameter group uses `reserved-memory` to set aside memory for Redis overhead, before you begin scaling be sure that you have a custom parameter group that reserves the correct amount of memory for your new node type. Alternatively, you can modify a custom parameter group so that it uses `reserved-memory-percent` and use that parameter group for your new cluster.

If you're using `reserved-memory-percent`, doing this is not necessary.

For more information, see Managing Reserved Memory (p. 466).
Scaling Up Single-Node Clusters for Redis (Cluster Mode Disabled) (Console)

The following procedure describes how to scale up a single-node Redis cluster using the ElastiCache Management Console. During this process, your Redis cluster will continue to serve requests with minimal downtime.

**To scale up a single-node Redis cluster (console)**

2. From the navigation pane, choose **Redis**.
3. From the list of clusters, choose the cluster you want to scale up (it must be running the Redis engine, not the Clustered Redis engine).
4. Choose **Modify**.
5. In the **Modify Cluster** wizard:
   a. Choose the node type you want to scale to from the **Node type** list.
   b. If you’re using reserved-memory to manage your memory, from the **Parameter Group** list, choose the custom parameter group that reserves the correct amount of memory for your new node type.
6. If you want to perform the scale up process right away, choose the **Apply immediately** box. If the **Apply immediately** box is not chosen, the scale-up process is performed during this cluster’s next maintenance window.
7. Choose **Modify**.
   a. If you chose **Apply immediately** in the previous step, the cluster’s status changes to **modifying**.
   b. When the status changes to **available**, the modification is complete and you can begin using the new cluster.

Scaling Up Single-Node Redis Cache Clusters (AWS CLI)

The following procedure describes how to scale up a single-node Redis cache cluster using the AWS CLI. During this process, your Redis cluster will continue to serve requests with minimal downtime.

**To scale up a single-node Redis cache cluster (AWS CLI)**

1. Determine the node types you can scale up to by running the AWS CLI `list-allowed-node-type-modifications` command with the following parameter.
   
   - `--cache-cluster-id`

   For Linux, macOS, or Unix:

   ```sh
   aws elasticache list-allowed-node-type-modifications \
   --cache-cluster-id my-cache-cluster-id
   ```

   For Windows:

   ```sh
   aws elasticache list-allowed-node-type-modifications ^
   --cache-cluster-id my-cache-cluster-id
   ```

   Output from the above command looks something like this (JSON format).

   ```json
   {}
   ```
"ScaleUpModifications": [ 
  "cache.m3.2xlarge",
  "cache.m3.large",
  "cache.m3.xlarge",
  "cache.m4.10xlarge",
  "cache.m4.2xlarge",
  "cache.m4.4xlarge",
  "cache.m4.large",
  "cache.m4.xlarge",
  "cache.r3.2xlarge",
  "cache.r3.4xlarge",
  "cache.r3.8xlarge",
  "cache.r3.large",
  "cache.r3.xlarge"
  ]

"ScaleDownModifications": [ 
  "cache.t2.micro",
  "cache.t2.small",
  "cache.t2.medium",
  "cache.t1.small",
  "cache.t1.xsmall"
  ]
}

For more information, see list-allowed-node-type-modifications in the AWS CLI Reference.

2. Modify your existing cache cluster specifying the cache cluster to scale up and the new, larger node type, using the AWS CLI modify-cache-cluster command and the following parameters.

- --cache-cluster-id – The name of the cache cluster you are scaling up or down.
- --cache-node-type – The new node type you want to scale the cache cluster. This value must be one of the node types returned by the list-allowed-node-type-modifications command in step 1.
- --cache-parameter-group-name – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
- --apply-immediately – Causes the scale-up process to be applied immediately. To postpone the scale-up process to the cluster's next maintenance window, use the --no-apply-immediately parameter.

For Linux, macOS, or Unix:

```bash
aws elasticsearch modify-cache-cluster \
  --cache-cluster-id my-redis-cache-cluster \
  --cache-node-type cache.m3.xlarge \
  --cache-parameter-group-name redis32-m2-xl \
  --apply-immediately
```

For Windows:

```bash
aws elasticsearch modify-cache-cluster ^
  --cache-cluster-id my-redis-cache-cluster ^
  --cache-node-type cache.m3.xlarge ^
  --cache-parameter-group-name redis32-m2-xl ^
  --apply-immediately
```

Output from the above command looks something like this (JSON format).
For more information, see modify-cache-cluster in the AWS CLI Reference.

3. If you used the --apply-immediately, check the status of the new cache cluster using the AWS CLI describe-cache-clusters command with the following parameter. When the status changes to available, you can begin using the new, larger cache cluster.

- --cache-cache-cluster-id– The name of your single-node Redis cache cluster. Use this parameter to describe a particular cache cluster rather than all cache clusters.

aws elasticache describe-cache-clusters --cache-cluster-id my-redis-cache-cluster

For more information, see describe-cache-clusters in the AWS CLI Reference.

Scaling Up Single-Node Redis Cache Clusters (ElastiCache API)

The following procedure describes how to scale up a single-node Redis cache cluster using the ElastiCache API. During this process, your Redis cluster will continue to serve requests with minimal downtime.

To scale up a single-node Redis cache cluster (ElastiCache API)

1. Determine the node types you can scale up to by running the ElastiCache API ListAllowedNodeTypeModifications action with the following parameter.

- CacheClusterId– The name of the single-node Redis cache cluster you want to scale up.

https://elasticache.us-west-2.amazonaws.com/?Action=ListAllowedNodeTypeModifications
For more information, see ListAllowedNodeTypeModifications in the Amazon ElastiCache API Reference.

2. Modify your existing cache cluster specifying the cache cluster to scale up and the new, larger node type, using the ModifyCacheCluster ElastiCache API action and the following parameters.

- **CacheClusterId** – The name of the cache cluster you are scaling up.
- **CacheNodeType** – The new, larger node type you want to scale the cache cluster up to. This value must be one of the node types returned by the ListAllowedNodeTypeModifications action in step 1.
- **CacheParameterGroupName** – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
- **ApplyImmediately** – Set to true to cause the scale-up process to be performed immediately. To postpone the scale-up process to the cluster's next maintenance window, use ApplyImmediately=false.

```plaintext
https://elasticache.us-west-2.amazonaws.com/
?Action=ModifyCacheCluster
&ApplyImmediately=true
&CacheClusterId=MyRedisCacheCluster
&CacheNodeType=cache.m3.xlarge
&CacheParameterGroupName=redis32-m2-xl
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>
```

For more information, see ModifyCacheCluster in the Amazon ElastiCache API Reference.

3. If you used ApplyImmediately=true, check the status of the new cache cluster using the ElastiCache API DescribeCacheClusters action with the following parameter. When the status changes to available, you can begin using the new, larger cache cluster.

- **CacheClusterId** – The name of your single-node Redis cache cluster. Use this parameter to describe a particular cache cluster rather than all cache clusters.

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

For more information, see DescribeCacheClusters in the Amazon ElastiCache API Reference.
Scaling Down Single-Node Redis Clusters

The following sections walk you through how to scale a single-node Redis cluster down to a smaller node type. Ensuring that the new, smaller node type is large enough to accommodate all the data and Redis overhead is important to the long-term success of your new Redis cluster. For more information, see Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464).

Topics
- Scaling Down a Single-Node Redis Cluster (Console) (p. 259)
- Scaling Down Single-Node Redis Cache Clusters (AWS CLI) (p. 260)
- Scaling Down Single-Node Redis Cache Clusters (ElastiCache API) (p. 262)

Scaling Down a Single-Node Redis Cluster (Console)

The following procedure walks you through scaling your single-node Redis cluster down to a smaller node type using the ElastiCache console.

Important
If your parameter group uses reserved-memory to set aside memory for Redis overhead, before you begin scaling be sure that you have a custom parameter group that reserves the correct amount of memory for your new node type. Alternatively, you can modify a custom parameter group so that it uses reserved-memory-percent and use that parameter group for your new cluster.
If you’re using reserved-memory-percent, doing this is not necessary.
For more information, see Managing Reserved Memory (p. 466).

To scale down your single-node Redis cluster (console)

1. Ensure that the smaller node type is adequate for your data and overhead needs.
2. If your parameter group uses reserved-memory to set aside memory for Redis overhead, ensure that you have a custom parameter group that sets aside the correct amount of memory for your new node type.
   Alternatively, you can modify your custom parameter group to use reserved-memory-percent.
   For more information, see Managing Reserved Memory (p. 466).
4. From the list of clusters, choose the cluster you want to scale down. This cluster must be running the Redis engine and not the Clustered Redis engine.
5. Choose Modify.
6. In the Modify Cluster wizard:
   a. Choose the node type you want to scale down to from the Node type list.
   b. If you’re using reserved-memory to manage your memory, from the Parameter Group list, choose the custom parameter group that reserves the correct amount of memory for your new node type.
7. If you want to perform the scale-down process right away, choose the Apply immediately check box. If the Apply immediately check box is left not chosen, the scale-down process is performed during this cluster’s next maintenance window.
8. Choose Modify.
9. When the cluster’s status changes from modifying to available, your cluster has scaled to the new node type. There is no need to update the endpoints in your application.
Scaling Down Single-Node Redis Cache Clusters (AWS CLI)

The following procedure describes how to scale down a single-node Redis cache cluster using the AWS CLI.

**To scale down a single-node Redis cache cluster (AWS CLI)**

1. Determine the node types you can scale down to by running the AWS CLI `list-allowed-node-type-modifications` command with the following parameter.

   • `--cache-cluster-id`

   For Linux, macOS, or Unix:

   ```sh
   aws elasticache list-allowed-node-type-modifications --cache-cluster-id my-cache-cluster-id
   ```

   For Windows:

   ```sh
   aws elasticache list-allowed-node-type-modifications --cache-cluster-id my-cache-cluster-id
   ```

   Output from the above command looks something like this (JSON format).

   ```json
   {
     "ScaleUpModifications": [
       "cache.m3.2xlarge",
       "cache.m3.large",
       "cache.m3.xlarge",
       "cache.m4.10xlarge",
       "cache.m4.2xlarge",
       "cache.m4.4xlarge",
       "cache.m4.large",
       "cache.m4.xlarge",
       "cache.r3.2xlarge",
       "cache.r3.4xlarge",
       "cache.r3.8xlarge",
       "cache.r3.large",
       "cache.r3.xlarge"
     ],
     "ScaleDownModifications": [
       "cache.t2.micro",
       "cache.t2.small",
       "cache.t2.medium",
       "cache.t1.medium",
       "cache.t1.small"
     ]
   }
   ```

   For more information, see `list-allowed-node-type-modifications` in the [AWS CLI Reference](https://docs.aws.amazon.com/cli/latest/reference/elasticache/list-allowed-node-type-modifications).

2. Modify your existing cache cluster specifying the cache cluster to scale down and the new, smaller node type, using the AWS CLI `modify-cache-cluster` command and the following parameters.

   • `--cache-cluster-id` – The name of the cache cluster you are scaling down.

   • `--cache-node-type` – The new node type you want to scale the cache cluster. This value must be one of the node types returned by the `list-allowed-node-type-modifications` command in step 1.
• `--cache-parameter-group-name` – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.

• `--apply-immediately` – Causes the scale-down process to be applied immediately. To postpone the scale-up process to the cluster's next maintenance window, use the `--no-apply-immediately` parameter.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-cache-cluster \
  --cache-cluster-id my-redis-cache-cluster \
  --cache-node-type cache.m3.xlarge \
  --cache-parameter-group-name redis32-m2-xl \
  --apply-immediately
```

For Windows:

```bash
aws elasticache modify-cache-cluster ^
  --cache-cluster-id my-redis-cache-cluster ^
  --cache-node-type cache.m3.xlarge ^
  --cache-parameter-group-name redis32-m2-xl ^
  --apply-immediately
```

Output from the above command looks something like this (JSON format).

```json
{
  "CacheCluster": {
    "Engine": "redis",
    "CacheParameterGroup": {
      "CacheNodeIdsToReboot": [],
      "CacheParameterGroupName": "default.redis3.2",
      "ParameterApplyStatus": "in-sync"
    },
    "SnapshotRetentionLimit": 1,
    "CacheClusterId": "my-redis-cache-cluster",
    "CacheSecurityGroups": [],
    "NumCacheNodes": 1,
    "SnapshotWindow": "00:00-01:00",
    "CacheClusterCreateTime": "2017-02-21T22:34:09.645Z",
    "AutoMinorVersionUpgrade": true,
    "CacheClusterStatus": "modifying",
    "PreferredAvailabilityZone": "us-west-2a",
home#client-download:",
    "CacheSubnetGroupName": "default",
    "EngineVersion": "3.2.4",
    "PendingModifiedValues": {
      "CacheNodeType": "cache.m3.2xlarge"
    },
    "PreferredMaintenanceWindow": "tue:11:30-tue:12:30",
    "CacheNodeType": "cache.m3.medium"
  }
}
```

For more information, see `modify-cache-cluster` in the AWS CLI Reference.
3. If you used the `--apply-immediately` option, check the status of the new cache cluster using the AWS CLI `describe-cache-clusters` command with the following parameter. When the status changes to `available`, you can begin using the new, larger cache cluster.

   • `--cache-cache-cluster-id` - The name of your single-node Redis cache cluster. Use this parameter to describe a particular cache cluster rather than all cache clusters.

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

   For more information, see `describe-cache-clusters` in the AWS CLI Reference.

## Scaling Down Single-Node Redis Cache Clusters (ElastiCache API)

The following procedure describes how to scale up/down a single-node Redis cache cluster using the ElastiCache API.

### To scale down a single-node Redis cache cluster (ElastiCache API)

1. Determine the node types you can scale down to by running the ElastiCache API `ListAllowedNodeTypeModifications` action with the following parameter.

   • `CacheClusterId` - The name of the single-node Redis cache cluster you want to scale down.

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

   For more information, see `ListAllowedNodeTypeModifications` in the Amazon ElastiCache API Reference.

2. Modify your existing cache cluster specifying the cache cluster to scale up and the new, larger node type, using the `ModifyCacheCluster` ElastiCache API action and the following parameters.

   • `CacheClusterId` - The name of the cache cluster you are scaling down.

   • `CacheNodeType` - The new, larger node type you want to scale the cache cluster down to. This value must be one of the node types returned by the `ListAllowedNodeTypeModifications` action in step 1.

   • `CacheParameterGroupName` - [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.

   • `ApplyImmediately` - Set to `true` to cause the scale-down process to be performed immediately. To postpone the scale-up process to the cluster's next maintenance window, use `ApplyImmediately=false`.

   ```url
   https://elasticache.us-west-2.amazonaws.com/
   ?Action=ModifyCacheCluster
   &ApplyImmediately=true
   &CacheClusterId=MyRedisCacheCluster
   &CacheNodeType=cache.m3.xlarge
   ```

API Version 2015-02-02

262
For more information, see ModifyCacheCluster in the Amazon ElastiCache API Reference.

3. If you used ApplyImmediately=true, check the status of the new cache cluster using the ElastiCache API DescribeCacheClusters action with the following parameter. When the status changes to available, you can begin using the new, smaller cache cluster.

- CacheClusterId – The name of your single-node Redis cache cluster. Use this parameter to describe a particular cache cluster rather than all cache clusters.

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

For more information, see DescribeCacheClusters in the Amazon ElastiCache API Reference.
Scaling Redis (Cluster Mode Disabled) Clusters with Replica Nodes

A Redis cluster with replica nodes (called replication group in the API/CLI) provides high availability via replication that has Multi-AZ with automatic failover enabled. A cluster with replica nodes is a logical collection of up to six Redis clusters where one node, the Primary, is able to serve both read and write requests. All the other nodes in the cluster are read-only replicas of the Primary. Data written to the Primary is asynchronously replicated to all the read replicas in the cluster. Because Redis (cluster mode disabled) does not support partitioning your data across multiple clusters, each node in a Redis (cluster mode disabled) replication group contains the entire cache dataset. Redis (cluster mode enabled) clusters support partitioning your data across up to 90 shards.

To change the data capacity of your cluster you must scale it up to a larger node type, or down to a smaller node type.

To change the read capacity of your cluster, add more read replicas, up to a maximum of 5, or remove read replicas.

The ElastiCache scaling up process is designed to make a best effort to retain your existing data and requires successful Redis replication. For Redis clusters with replicas, we recommend that sufficient memory be made available to Redis.

Related Topics

- High Availability Using Replication Groups (p. 142)
- Replication: Redis (Cluster Mode Disabled) vs. Redis (Cluster Mode Enabled) (p. 145)
- Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148)
- Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464)

Topics

- Scaling Up Redis Clusters with Replicas (p. 265)
- Scaling Down Redis Clusters with Replicas (p. 271)
- Increasing Read Capacity (p. 277)
- Decreasing Read Capacity (p. 278)
Amazon ElastiCache for Redis
ElastiCache for Redis User Guide
Scaling Clusters for Redis (Cluster Mode Disabled)

Scaling Up Redis Clusters with Replicas

Amazon ElastiCache provides console, CLI, and API support for scaling your Redis (cluster mode disabled) replication group up.

When the scale-up process is initiated, ElastiCache does the following:

1. Launches a replication group using the new node type.
2. Copies all the data from the current primary node to the new primary node.
3. Syncs the new read replicas with the new primary node.
4. Updates the DNS entries so they point to the new nodes. Because of this you don't have to update the endpoints in your application. For Redis 5.0.5 and above, you can scale auto failover enabled clusters while the cluster continues to stay online and serve incoming requests. On version 5.0.4 and below, you may notice a brief interruption of reads and writes on previous versions from the primary node while the DNS entry is updated.
5. Deletes the old nodes (CLI/API: replication group).

How long this process takes is dependent upon your node type and how much data is in your cluster.

As shown in the following table, your Redis scale-up operation is blocked if you have an engine upgrade scheduled for the cluster's next maintenance window.

Blocked Redis operations

<table>
<thead>
<tr>
<th>Pending Operations</th>
<th>Blocked Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale up</td>
<td>Immediate engine upgrade</td>
</tr>
<tr>
<td>Engine upgrade</td>
<td>Immediate scale up</td>
</tr>
<tr>
<td>Scale up and engine upgrade</td>
<td>Immediate scale up</td>
</tr>
<tr>
<td></td>
<td>Immediate engine upgrade</td>
</tr>
</tbody>
</table>

If you have a pending operation that is blocking you, you can do one of the following.

- Schedule your Redis scale-up operation for the next maintenance window by clearing the **Apply immediately** check box (CLI use: --no-apply-immediately, API use: ApplyImmediately=false).
- Wait until your next maintenance window (or after) to perform your Redis scale-up operation.
- Add the Redis engine upgrade to this cache cluster modification with the **Apply Immediately** check box chosen (CLI use: --apply-immediately, API use: ApplyImmediately=true). This unblocks your scale-up operation by causing the engine upgrade to be performed immediately.

The following sections describe how to scale your Redis cluster with replicas up using the ElastiCache console, the AWS CLI, and the ElastiCache API.

**Important**

If your parameter group uses reserved-memory to set aside memory for Redis overhead, before you begin scaling be sure that you have a custom parameter group that reserves the correct amount of memory for your new node type. Alternatively, you can modify a custom parameter group so that it uses reserved-memory-percent and use that parameter group for your new cluster.

If you're using reserved-memory-percent, doing this is not necessary.
For more information, see Managing Reserved Memory (p. 466).
Scaling Up a Redis Cluster with Replicas (Console)

The amount of time it takes to scale up to a larger node type varies, depending upon the node type and the amount of data in your current cluster.

The following process scales your cluster with replicas from its current node type to a new, larger node type using the ElastiCache console. During this process, there may be a brief interruption of reads and writes for other versions from the primary node while the DNS entry is updated. you might see less than 1 second downtime for nodes running on 5.0.5 versions and a few seconds for older versions.

To scale up Redis cluster with replicas (console)

2. From the navigation pane, choose Redis.
3. From the list of clusters, choose the cluster you want to scale up. This cluster must be running the Redis engine and not the Clustered Redis engine.
4. Choose Modify.
5. In the Modify Cluster wizard:
   a. Choose the node type you want to scale to from the Node type list. Note that not all node types are available to scale down to.
   b. If you’re using reserved-memory to manage your memory, from the Parameter Group list, choose the custom parameter group that reserves the correct amount of memory for your new node type.
6. If you want to perform the scale-up process right away, choose the Apply immediately check box. If the Apply immediately check box is left not chosen, the scale-up process is performed during this cluster’s next maintenance window.
7. Choose Modify.
8. When the cluster’s status changes from modifying to available, your cluster has scaled to the new node type. There is no need to update the endpoints in your application.

Scaling Up a Redis Replication Group (AWS CLI)

The following process scales your replication group from its current node type to a new, larger node type using the AWS CLI. During this process, until the status changes from modifying to available, all reads and writes between your application and the primary cache cluster are blocked.

The amount of time it takes to scale up to a larger node type varies, depending upon your node type and the amount of data in your current cache cluster.

To scale up a Redis Replication Group (AWS CLI)

1. Determine which node types you can scale up to by running the AWS CLI list-allowed-node-type-modifications command with the following parameter.

   • --replication-group-id—the name of the replication group. Use this parameter to describe a particular replication group rather than all replication groups.

   For Linux, macOS, or Unix:
   
   ```bash
   aws elasticache list-allowed-node-type-modifications \ 
   --replication-group-id my-repl-group
   ```

   For Windows:

   ```bash
   aws elasticache list-allowed-node-type-modifications \ 
   --replication-group-id my-repl-group
   ```
aws elasticache list-allowed-node-type-modifications ^
   --replication-group-id my-repl-group

Output from this operation looks something like this (JSON format).

```
{
   "ScaleUpModifications": [
      "cache.m3.2xlarge",
      "cache.m3.large",
      "cache.m3.xlarge",
      "cache.m4.10xlarge",
      "cache.m4.2xlarge",
      "cache.m4.4xlarge",
      "cache.m4.large",
      "cache.m4.xlarge",
      "cache.r3.2xlarge",
      "cache.r3.4xlarge",
      "cache.r3.8xlarge",
      "cache.r3.large",
      "cache.r3.xlarge"
   ]
}
```

For more information, see list-allowed-node-type-modifications in the AWS CLI Reference.

2. Scale your current replication group up to the new node type using the AWS CLI modify-replication-group command with the following parameters.

   • `--replication-group-id` – the name of the replication group.
   • `--cache-node-type` – the new, larger node type of the cache clusters in this replication group. This value must be one of the instance types returned by the list-allowed-node-type-modifications command in step 1.
   • `--cache-parameter-group-name` – [Optional] Use this parameter if you are using reserved-memory to manage your cluster’s reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
   • `--apply-immediately` – Causes the scale-up process to be applied immediately. To postpone the scale-up operation to the next maintenance window, use `--no-apply-immediately`.

For Linux, macOS, or Unix:

```
aws elasticache modify-replication-group \
   --replication-group-id my-repl-group \
   --cache-node-type cache.m3.xlarge \
   --cache-parameter-group-name redis32-m3-2xl \
   --apply-immediately
```

For Windows:

```
aws elasticache modify-replication-group ^
   --replication-group-id my-repl-group ^
   --cache-node-type cache.m3.xlarge ^
   --cache-parameter-group-name redis32-m3-2xl ^
   --apply-immediately
```

Output from this command looks something like this (JSON format).
3. If you used the `--apply-immediately` parameter, monitor the status of the replication group using the AWS CLI `describe-replication-group` command with the following parameter. While the status is still in modifying, you might see less than 1 second downtime for nodes running on 5.0.5 versions and a brief interruption of reads and writes for older versions from the primary node while the DNS entry is updated.

- `--replication-group-id` – the name of the replication group. Use this parameter to describe a particular replication group rather than all replication groups.

For Linux, macOS, or Unix:

```
aws elasticache describe-replication-group \
```
Scaling Clusters for Redis (Cluster Mode Disabled)

--replication-group-id my-replication-group

For Windows:

aws elasticache describe-replication-groups
  --replication-group-id my-replication-group

For more information, see describe-replication-groups in the AWS CLI Reference.

Scaling Up a Redis Replication Group (ElastiCache API)

The following process scales your replication group from its current node type to a new, larger node type using the ElastiCache API. During this process, until the status changes from modifying to available, all reads and writes between your application and the primary cache cluster are blocked. However, reads from the read replica cache clusters continue uninterrupted.

The amount of time it takes to scale up to a larger node type varies, depending upon your node type and the amount of data in your current cache cluster.

To scale up a Redis Replication Group (ElastiCache API)

1. Determine which node types you can scale up to using the ElastiCache API ListAllowedNodeTypeModifications action with the following parameter.
   - ReplicationGroupId – the name of the replication group. Use this parameter to describe a specific replication group rather than all replication groups.

   https://elasticache.us-west-2.amazonaws.com/
   ?Action=ListAllowedNodeTypeModifications
   &ReplicationGroupId=MyReplGroup
   &Version=2015-02-02
   &SignatureVersion=4
   &SignatureMethod=HmacSHA256
   &Timestamp=20150202T192317Z
   &X-Amz-Credential=<credential>

   For more information, see ListAllowedNodeTypeModifications in the Amazon ElastiCache API Reference.

2. Scale your current replication group up to the new node type using the ModifyReplicationGroup ElastiCache API action and with the following parameters.
   - ReplicationGroupId – the name of the replication group.
   - CacheNodeType – the new, larger node type of the cache clusters in this replication group. This value must be one of the instance types returned by the ListAllowedNodeTypeModifications action in step 1.
   - CacheParameterGroupName – [Optional] Use this parameter if you are using reserved-memory to manage your cluster’s reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
   - ApplyImmediately – Set to true to causes the scale-up process to be applied immediately. To postpone the scale-up process to the next maintenance window, use ApplyImmediately=false.

   https://elasticache.us-west-2.amazonaws.com/
   ?Action=ModifyReplicationGroup
&ApplyImmediately=true
&CacheNodeType=cache.m3.2xlarge
&CacheParameterGroupName=redis32-m3-2xl
&ReplicationGroupId=myReplGroup
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20141201T220302Z
&Version=2014-12-01
&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 ModifyReplicationGroup in the Amazon ElastiCache API Reference.

3. If you used ApplyImmediately=true, monitor the status of the replication group using the ElastiCache API DescribeReplicationGroups action with the following parameters. When the status changes from modifying to available, you can begin writing to your new, scaled up replication group.

• ReplicationGroupId – the name of the replication group. Use this parameter to describe a particular replication group rather than all replication groups.

https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeReplicationGroups
&ReplicationGroupId=MyReplGroup
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>

For more information, see DescribeReplicationGroups in the Amazon ElastiCache API Reference.
Scaling Down Redis Clusters with Replicas

The following sections walk you through how to scale a Redis (cluster mode disabled) cache cluster with replica nodes down to a smaller node type. Ensuring that the new, smaller node type is large enough to accommodate all the data and overhead is very important to success. For more information, see Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464).

Important
If your parameter group uses reserved-memory to set aside memory for Redis overhead, before you begin scaling be sure that you have a custom parameter group that reserves the correct amount of memory for your new node type. Alternatively, you can modify a custom parameter group so that it uses reserved-memory-percent and use that parameter group for your new cluster.
If you're using reserved-memory-percent, doing this is not necessary.
For more information, see Managing Reserved Memory (p. 466).

Topics
• Scaling Down a Redis Cluster with Replicas (Console) (p. 271)
• Scaling Down a Redis Replication Group (AWS CLI) (p. 266)
• Scaling Down a Redis Replication Group (ElastiCache API) (p. 274)

Scaling Down a Redis Cluster with Replicas (Console)

The following process scales your Redis cluster with replica nodes to a smaller node type using the ElastiCache console.

To scale down a Redis cluster with replica nodes (console)

1. Ensure that the smaller node type is adequate for your data and overhead needs.
2. If your parameter group uses reserved-memory to set aside memory for Redis overhead, ensure that you have a custom parameter group to set aside the correct amount of memory for your new node type.

   Alternatively, you can modify your custom parameter group to use reserved-memory-percent. For more information, see Managing Reserved Memory (p. 466).

4. From the list of clusters, choose the cluster you want to scale down. This cluster must be running the Redis engine and not the Clustered Redis engine.
5. Choose Modify.
6. In the Modify Cluster wizard:
   a. Choose the node type you want to scale down to from the Node type list.
   b. If you're using reserved-memory to manage your memory, from the Parameter Group list, choose the custom parameter group that reserves the correct amount of memory for your new node type.

7. If you want to perform the scale-down process right away, choose the Apply immediately check box. If the Apply immediately check box is left not chosen, the scale-down process is performed during this cluster's next maintenance window.
8. Choose Modify.
9. When the cluster’s status changes from modifying to available, your cluster has scaled to the new node type. There is no need to update the endpoints in your application.
Scaling Down a Redis Replication Group (AWS CLI)

The following process scales your replication group from its current node type to a new, smaller node type using the AWS CLI. During this process, until the status changes from modifying to available, all reads and writes between your application and the primary cache cluster are blocked.

The amount of time it takes to scale down to a smaller node type varies, depending upon your node type and the amount of data in your current cache cluster.

To scale down a Redis Replication Group (AWS CLI)

1. Determine which node types you can scale down to by running the AWS CLI list-allowed-node-type-modifications command with the following parameter.

   • `--replication-group-id` – the name of the replication group. Use this parameter to describe a particular replication group rather than all replication groups.

   For Linux, macOS, or Unix:

   ```
   aws elasticache list-allowed-node-type-modifications \ 
   --replication-group-id my-repl-group
   ```

   For Windows:

   ```
   aws elasticache list-allowed-node-type-modifications ^
   --replication-group-id my-repl-group
   ```

   Output from this operation looks something like this (JSON format).

   ```json
   {
   "ScaleDownModifications": [
   "cache.m3.2xlarge",
   "cache.m3.large",
   "cache.m3.xlarge",
   "cache.m4.10xlarge",
   "cache.m4.2xlarge",
   "cache.m4.4xlarge",
   "cache.m4.large",
   "cache.m4.xlarge",
   "cache.r3.2xlarge",
   "cache.r3.4xlarge",
   "cache.r3.8xlarge",
   "cache.r3.large",
   "cache.r3.xlarge"
   ]
   }
   ```

   For more information, see list-allowed-node-type-modifications in the AWS CLI Reference.

2. Scale your current replication group up to the new node type using the AWS CLI modify-replication-group command with the following parameters.

   • `--replication-group-id` – the name of the replication group.

   • `--cache-node-type` – the new, larger node type of the cache clusters in this replication group. This value must be one of the instance types returned by the list-allowed-node-type-modifications command in step 1.

   • `--cache-parameter-group-name` – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group
that reserves the correct amount of memory for your new node type. If you are using `reserved-memory-percent` you can omit this parameter.

- `--apply-immediately` – Causes the scale-up process to be applied immediately. To postpone the scale-up operation to the next maintenance window, use `--no-apply-immediately`.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group
  --replication-group-id my-repl-group
  --cache-node-type cache.t2.small
  --cache-parameter-group-name redis32-m3-2xl
  --apply-immediately
```

For Windows:

```bash
aws elasticache modify-replication-group
  --replication-group-id my-repl-group
  --cache-node-type cache.t2.small
  --cache-parameter-group-name redis32-m3-2xl
  --apply-immediately
```

Output from this command looks something like this (JSON format).

```json
{"ReplicationGroup": {
  "Status": "available",
  "Description": "Some description",
  "NodeGroups": [
    {
      "Status": "available",
      "NodeGroupMembers": [
        {
          "CurrentRole": "primary",
          "PreferredAvailabilityZone": "us-west-2b",
          "CacheNodeId": "0001",
          "ReadEndpoint": {
            "Port": 6379,
            "Address": "my-repl-group-001.8fdx4s.0001.usw2.cache.amazonaws.com"
          },
          "CacheClusterId": "my-repl-group-001"
        },
        {
          "CurrentRole": "replica",
          "PreferredAvailabilityZone": "us-west-2c",
          "CacheNodeId": "0001",
          "ReadEndpoint": {
            "Port": 6379,
            "Address": "my-repl-group-002.8fdx4s.0001.usw2.cache.amazonaws.com"
          },
          "CacheClusterId": "my-repl-group-002"
        }
      ],
      "NodeGroupId": "0001",
      "PrimaryEndpoint": {
        "Port": 6379,
        "Address": "my-repl-group.8fdx4s.ng.0001.usw2.cache.amazonaws.com"
      }
    }
  ]
}
```
3. If you used the `--apply-immediately` parameter, monitor the status of the replication group using the AWS CLI `describe-replication-group` command with the following parameter. When the status changes from `modifying` to `available`, you can begin writing to your new, scaled down replication group.

```bash
For Linux, macOS, or Unix:
aws elasticache describe-replication-group \
   --replication-group-id my-replication-group
```

```bash
For Windows:
aws elasticache describe-replication-groups ^
   --replication-group-id my-replication-group
```

For more information, see `describe-replication-groups` in the AWS CLI Reference.

**Scaling Down a Redis Replication Group (ElastiCache API)**

The following process scales your replication group from its current node type to a new, smaller node type using the ElastiCache API. During this process, until the status changes from `modifying` to `available`, all reads and writes between your application and the primary cache cluster are blocked. However, reads from the read replica cache clusters continue uninterrupted.

The amount of time it takes to scale down to a smaller node type varies, depending upon your node type and the amount of data in your current cache cluster.

**To scale down a Redis Replication Group (ElastiCache API)**

1. Determine which node types you can scale down to using the ElastiCache API `ListAllowedNodeTypeModifications` action with the following parameter.

```bash
https://elasticache.us-west-2.amazonaws.com/
?Action=ListAllowedNodeTypeModifications
&ReplicationGroupId=MyReplGroup
```

For more information, see `modify-replication-group` in the AWS CLI Reference.
For more information, see ListAllowedNodeTypeModifications in the Amazon ElastiCache API Reference.

2. Scale your current replication group up to the new node type using the ModifyReplicationGroup ElastiCache API action and with the following parameters.

- ReplicationGroupId – the name of the replication group.
- CacheNodeType – the new, larger node type of the cache clusters in this replication group. This value must be one of the instance types returned by the ListAllowedNodeTypeModifications action in step 1.
- CacheParameterGroupName – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
- ApplyImmediately – Set to true to causes the scale-up process to be applied immediately. To postpone the scale-down process to the next maintenance window, use ApplyImmediately=false.

For more information, see ModifyReplicationGroup in the Amazon ElastiCache API Reference.

3. If you used ApplyImmediately=true, monitor the status of the replication group using the ElastiCache API DescribeReplicationGroups action with the following parameters. When the status changes from modifying to available, you can begin writing to your new, scaled down replication group.

- ReplicationGroupId – the name of the replication group. Use this parameter to describe a particular replication group rather than all replication groups.
&X-Amz-Credential=<credential>

For more information, see DescribeReplicationGroups in the Amazon ElastiCache API Reference.
Increasing Read Capacity

To increase read capacity, add read replicas (up to a maximum of five) to your Redis replication group.

You can scale your Redis cluster's read capacity using the ElastiCache console, the AWS CLI, or the ElastiCache API. For more information, see Adding a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 200).
Decreasing Read Capacity

To decrease read capacity, delete one or more read replicas from your Redis cluster with replicas (called replication group in the API/CLI). If the cluster is Multi-AZ with automatic failover enabled, you cannot delete the last read replica without first disabling Multi-AZ with automatic failover. For more information, see Modifying a Replication Group (p. 188).

For more information, see Deleting a Read Replica, for Redis (Cluster Mode Disabled) Replication Groups (p. 202).
Scaling Clusters in Redis (Cluster Mode Enabled)

As demand on your clusters changes, you might decide to improve performance or reduce costs by changing the number of shards in your Redis (cluster mode enabled) cluster. We recommend using online horizontal scaling to do so, because it allows your cluster to continue serving requests during the scaling process.

Conditions under which you might decide to rescale your cluster include the following:

- **Memory pressure:**
  
  If the nodes in your cluster are under memory pressure, you might decide to scale out so that you have more resources to better store data and serve requests.

  You can determine whether your nodes are under memory pressure by monitoring the following metrics: `FreeableMemory`, `SwapUsage`, and `BytesUseForCache`.

- **CPU or network bottleneck:**
  
  If latency/throughput issues are plaguing your cluster, you might need to scale out to resolve the issues.

  You can monitor your latency and throughput levels by monitoring the following metrics: `CPUUtilization`, `NetworkBytesIn`, `NetworkBytesOut`, `CurrConnections`, and `NewConnections`.

- **Your cluster is over-scaled:**
  
  Current demand on your cluster is such that scaling in doesn't hurt performance and reduces your costs.

  You can monitor your cluster's use to determine whether or not you can safely scale in using the following metrics: `FreeableMemory`, `SwapUsage`, `BytesUseForCache`, `CPUUtilization`, `NetworkBytesIn`, `NetworkBytesOut`, `CurrConnections`, and `NewConnections`.

**Performance Impact of Scaling**

When you scale using the offline process, your cluster is offline for a significant portion of the process and thus unable to serve requests. When you scale using the online method, because scaling is a compute-intensive operation, there is some degradation in performance, nevertheless, your cluster continues to serve requests throughout the scaling operation. How much degradation you experience depends upon your normal CPU utilization and your data.

There are two ways to scale your Redis (cluster mode enabled) cluster; horizontal and vertical scaling.

- **Horizontal scaling** allows you to change the number of node groups (shards) in the replication group by adding or removing node groups (shards). The online resharding process allows scaling in/out while the cluster continues serving incoming requests.

  Configure the slots in your new cluster differently than they were in the old cluster. Offline method only.

- **Vertical Scaling** - Change the node type to resize the cluster. The online vertical scaling allows scaling up/down while the cluster continues serving incoming requests.

If you are reducing the size and memory capacity of the cluster, by either scaling in or scaling down, ensure that the new configuration has sufficient memory for your data and Redis overhead.

For more information, see Choosing Your Node Size (p. 76).

Contents
Offline Resharding and Shard Rebalancing for Redis (cluster mode enabled)

The main advantage you get from offline shard reconfiguration is that you can do more than merely add or remove shards from your replication group. When you reshard offline, in addition to changing the number of shards in your replication group, you can do the following:

- Change the node type of your replication group.
- Specify the Availability Zone for each node in the replication group.
- Upgrade to a newer engine version.
- Specify the number of replica nodes in each shard independently.
- Specify the keyspace for each shard.

The main disadvantage of offline shard reconfiguration is that your cluster is offline beginning with the restore portion of the process and continuing until you update the endpoints in your application. The length of time that your cluster is offline varies with the amount of data in your cluster.

To reconfigure your shards Redis (cluster mode enabled) cluster offline

1. Create a manual backup of your existing Redis cluster. For more information, see Making Manual Backups (p. 218).
2. Create a new cluster by restoring from the backup. For more information, see Restoring From a Backup with Optional Cluster Resizing (p. 289).
Online Resharding and Shard Rebalancing for Redis (cluster mode enabled)

By using online resharding and shard rebalancing with Amazon ElastiCache for Redis version 3.2.10 or newer, you can scale your ElastiCache for Redis (cluster mode enabled) dynamically with no downtime. This approach means that your cluster can continue to serve requests even while scaling or rebalancing is in process.

You can do the following:

- **Scale out** – Increase read and write capacity by adding shards (node groups) to your Redis (cluster mode enabled) cluster (replication group).
  
  If you add one or more shards to your replication group, the number of nodes in each new shard is the same as the number of nodes in the smallest of the existing shards.

- **Scale in** – Reduce read and write capacity, and thereby costs, by removing shards from your Redis (cluster mode enabled) cluster.

- **Rebalance** – Move the keyspaces among the shards in your ElastiCache for Redis (cluster mode enabled) cluster so they are as equally distributed among the shards as possible.

You can’t do the following:

- **Configure shards independently:**
  
  - You can’t specify the number of nodes in each shard independently.
  - You can’t specify the keyspace for shards independently. To do this, you must use the offline process.

Currently, the following limitations apply to ElastiCache for Redis online resharding and rebalancing:

- These processes require Redis engine version 3.2.10 or newer. For information on upgrading your engine version, see Upgrading Engine Versions (p. 53).

- There are limitations with slots or keyspaces and large items:
  
  If any of the keys in a shard contain a large item, that key isn’t migrated to a new shard when scaling out or rebalancing. This functionality can result in unbalanced shards.

  If any of the keys in a shard contain a large item (items greater than 256 MB after serialization), that shard isn’t deleted when scaling in. This functionality can result in some shards not being deleted.

  - When scaling out, the number of nodes in any new shards equals the number of nodes in the smallest existing shard.
  - When scaling out, any tags that are common to all existing shards are copied to the new shards.

For more information, see Best Practices: Online Cluster Resizing (p. 472).

You can horizontally scale or rebalance your ElastiCache for Redis (cluster mode enabled) clusters using the AWS Management Console, the AWS CLI, and the ElastiCache API.

Adding Shards with Online Resharding

You can add shards to your Redis (cluster mode enabled) cluster using the AWS Management Console, AWS CLI, or ElastiCache API. When you add shards to a Redis (cluster mode enabled) cluster, any tags on the existing shards are copied over to the new shards.
Adding Shards (Console)

You can use the AWS Management Console to add one or more shards to your Redis (cluster mode enabled) cluster. The following procedure describes the process.

To add shards to your Redis (cluster mode enabled) cluster

2. From the navigation pane, choose Redis.
3. Locate and choose the name, not the box to the left of the cluster's name, of the Redis (cluster mode enabled) cluster that you want to add shards to.
   
   Tip
   Redis (cluster mode enabled) clusters have a value of 1 or greater in the Shards column.
4. Choose Add shard.

a. For Number of shards to be added, choose the number of shards you want added to this cluster.

b. For Availability zone(s), choose either No preference or Specify availability zones.

c. If you chose Specify availability zones, for each node in each shard, select the node's Availability Zone from the list of Availability Zones.

d. Choose Add.

Adding Shards (AWS CLI)

The following process describes how to reconfigure the shards in your Redis (cluster mode enabled) cluster by adding shards using the AWS CLI.

Use the following parameters with modify-replication-group-shard-configuration.

Parameters

- --apply-immediately – Required. Specifies the shard reconfiguration operation is to be started immediately.

- --replication-group-id – Required. Specifies which replication group (cluster) the shard reconfiguration operation is to be performed on.

- --node-group-count – Required. Specifies the number of shards (node groups) to exist when the operation is completed. When adding shards, the value of --node-group-count must be greater than the current number of shards.

   Optionally, you can specify the Availability Zone for each node in the replication group using --resharding-configuration.

- --resharding-configuration – Optional. A list of preferred Availability Zones for each node in each shard in the replication group. Use this parameter only if the value of --node-group-count is greater than the current number of shards. If this parameter is omitted when adding shards, Amazon ElastiCache selects the Availability Zones for the new nodes.
The following example reconfigures the keyspaces over four shards in the Redis (cluster mode enabled) cluster `my-cluster`. The example also specifies the Availability Zone for each node in each shard. The operation begins immediately.

**Example - Adding Shards**

For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group-shard-configuration
  --replication-group-id my-cluster
  --node-group-count 4
  --resharding-configuration
  "PreferredAvailabilityZones=us-east-2a,us-east-2c"
  "PreferredAvailabilityZones=us-east-2b,us-east-2a"
  "PreferredAvailabilityZones=us-east-2c,us-east-2d"
  "PreferredAvailabilityZones=us-east-2d,us-east-2c"
  --apply-immediately
```

For Windows:

```bash
aws elasticache modify-replication-group-shard-configuration ^
  --replication-group-id my-cluster ^
  --node-group-count 4 ^
  --resharding-configuration ^
  "PreferredAvailabilityZones=us-east-2a,us-east-2c" ^
  "PreferredAvailabilityZones=us-east-2b,us-east-2a" ^
  "PreferredAvailabilityZones=us-east-2c,us-east-2d" ^
  "PreferredAvailabilityZones=us-east-2d,us-east-2c" ^
  --apply-immediately
```

For more information, see `modify-replication-group-shard-configuration` in the AWS CLI documentation.

**Adding Shards (ElastiCache API)**

You can use the ElastiCache API to reconfigure the shards in your Redis (cluster mode enabled) cluster online by using the `ModifyReplicationGroupShardConfiguration` operation.

Use the following parameters with `ModifyReplicationGroupShardConfiguration`.

**Parameters**

- **ApplyImmediately**=true – Required. Specifies the shard reconfiguration operation is to be started immediately.
- **ReplicationGroupId** – Required. Specifies which replication group (cluster) the shard reconfiguration operation is to be performed on.
- **NodeGroupCount** – Required. Specifies the number of shards (node groups) to exist when the operation is completed. When adding shards, the value of `NodeGroupCount` must be greater than the current number of shards.

Optionally, you can specify the Availability Zone for each node in the replication group using `ReshardingConfiguration`.

- **ReshardingConfiguration** – Optional. A list of preferred Availability Zones for each node in each shard in the replication group. Use this parameter only if the value of `NodeGroupCount` is greater than the current number of shards. If this parameter is omitted when adding shards, Amazon ElastiCache selects the Availability Zones for the new nodes.

The following process describes how to reconfigure the shards in your Redis (cluster mode enabled) cluster by adding shards using the ElastiCache API.
Example - Adding Shards

The following example adds node groups to the Redis (cluster mode enabled) cluster my-cluster, so there are a total of four node groups when the operation completes. The example also specifies the Availability Zone for each node in each shard. The operation begins immediately.

https://elasticache.us-east-2.amazonaws.com/
    ?Action=ModifyReplicationGroupShardConfiguration
    &ApplyImmediately=true
    &NodeGroupCount=4
    &ReplicationGroupId=my-cluster
    &ReshardingConfiguration.1.PreferredAvailabilityZones.AvailabilityZone.1=us-east-2a
    &ReshardingConfiguration.1.PreferredAvailabilityZones.AvailabilityZone.2=us-east-2c
    &ReshardingConfiguration.2.PreferredAvailabilityZones.AvailabilityZone.1=us-east-2b
    &ReshardingConfiguration.2.PreferredAvailabilityZones.AvailabilityZone.2=us-east-2a
    &ReshardingConfiguration.3.PreferredAvailabilityZones.AvailabilityZone.1=us-east-2c
    &ReshardingConfiguration.3.PreferredAvailabilityZones.AvailabilityZone.2=us-east-2d
    &ReshardingConfiguration.4.PreferredAvailabilityZones.AvailabilityZone.1=us-east-2d
    &ReshardingConfiguration.4.PreferredAvailabilityZones.AvailabilityZone.2=us-east-2c
    &Version=2015-02-02
    &SignatureVersion=4
    &SignatureMethod=HmacSHA256
    &Timestamp=20171002T192317Z
    &X-Amz-Credential=<credential>

For more information, see ModifyReplicationGroupShardConfiguration in the ElastiCache API Reference.

Removing Shards with Online Resharding

You can remove shards from your Redis (cluster mode enabled) cluster using the AWS Management Console, AWS CLI, or ElastiCache API.

Topics

- Removing Shards (Console) (p. 284)
- Removing Shards (AWS CLI) (p. 285)
- Removing Shards (ElastiCache API) (p. 286)

Removing Shards (Console)

The following process describes how to reconfigure the shards in your Redis (cluster mode enabled) cluster by removing shards using the AWS Management Console.

Before removing node groups (shards) from your replication group, ElastiCache makes sure that all your data will fit in the remaining shards. If the data will fit, the specified shards are deleted from
the replication group as requested. If the data won't fit in the remaining node groups, the process is
terminated and the replication group is left with the same node group configuration as before the
request was made.

You can use the AWS Management Console to remove one or more shards from your Redis (cluster
mode enabled) cluster. You cannot remove all the shards in a replication group. Instead, you must delete
the replication group. For more information, see Deleting a Replication Group (p. 190). The following
procedure describes the process for deleting one or more shards.

**To remove shards from your Redis (cluster mode enabled) cluster**

2. From the navigation pane, choose Redis.
3. Locate and choose the name, not the box to the left of the cluster's name, of the Redis (cluster mode
   enabled) cluster you want to remove shards from.

   **Tip**
   Redis (cluster mode enabled) clusters have a value of 1 or greater in the Shards column.

4. From the list of shards, choose the box to the left of the name of each shard that you want to delete.
5. Choose Delete shard.

**Removing Shards (AWS CLI)**

The following process describes how to reconfigure the shards in your Redis (cluster mode enabled)
cluster by removing shards using the AWS CLI.

**Important**
Before removing node groups (shards) from your replication group, ElastiCache makes sure that
all your data will fit in the remaining shards. If the data will fit, the specified shards (--node-
groups-to-remove) are deleted from the replication group as requested and their keyspaces
mapped into the remaining shards. If the data will not fit in the remaining node groups, the
process is terminated and the replication group is left with the same node group configuration
as before the request was made.

You can use the AWS CLI to remove one or more shards from your Redis (cluster mode enabled) cluster.
You cannot remove all the shards in a replication group. Instead, you must delete the replication group.
For more information, see Deleting a Replication Group (p. 190).

Use the following parameters with modify-replication-group-shard-configuration.

**Parameters**

- **--apply-immediately** – Required. Specifies the shard reconfiguration operation is to be started
  immediately.
- **--replication-group-id** – Required. Specifies which replication group (cluster) the shard
  reconfiguration operation is to be performed on.
- **--node-group-count** – Required. Specifies the number of shards (node groups) to exist when the
  operation is completed. When removing shards, the value of --node-group-count must be less than
  the current number of shards.
- **--node-groups-to-remove** – Required when --node-group-count is less than the current
  number of node groups (shards). A list of shard (node group) IDs to remove from the replication group.

The following procedure describes the process for deleting one or more shards.
Example - Removing Shards

The following example removes two node groups from the Redis (cluster mode enabled) cluster `my-cluster`, so there are a total of two node groups when the operation completes. The keyspaces from the removed shards are distributed evenly over the remaining shards.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group-shard-configuration \
  --replication-group-id my-cluster \ 
  --node-group-count 2 \ 
  --node-groups-to-remove "0002" "0003" \ 
  --apply-immediately
```

For Windows:

```bash
aws elasticache modify-replication-group-shard-configuration ^ 
  --replication-group-id my-cluster ^ 
  --node-group-count 2 ^ 
  --node-groups-to-remove "0002" "0003" ^ 
  --apply-immediately
```

For more information, see `modify-replication-group-shard-configuration` in the AWS CLI documentation.

Removing Shards (ElastiCache API)

You can use the ElastiCache API to reconfigure the shards in your Redis (cluster mode enabled) cluster online by using the `ModifyReplicationGroupShardConfiguration` operation.

The following process describes how to reconfigure the shards in your Redis (cluster mode enabled) cluster by removing shards using the ElastiCache API.

**Important**

Before removing node groups (shards) from your replication group, ElastiCache makes sure that all your data will fit in the remaining shards. If the data will fit, the specified shards (`NodeGroupsToRemove`) are deleted from the replication group as requested and their keyspaces mapped into the remaining shards. If the data will not fit in the remaining node groups, the process is terminated and the replication group is left with the same node group configuration as before the request was made.

You can use the ElastiCache API to remove one or more shards from your Redis (cluster mode enabled) cluster. You cannot remove all the shards in a replication group. Instead, you must delete the replication group. For more information, see Deleting a Replication Group (p. 190).

Use the following parameters with `ModifyReplicationGroupShardConfiguration`.

**Parameters**

- **ApplyImmediately=true** – Required. Specifies the shard reconfiguration operation is to be started immediately.
- **ReplicationGroupId** – Required. Specifies which replication group (cluster) the shard reconfiguration operation is to be performed on.
- **NodeGroupCount** – Required. Specifies the number of shards (node groups) to exist when the operation is completed. When removing shards, the value of `NodeGroupCount` must be less than the current number of shards.
- **NodeGroupsToRemove** – Required when `--node-group-count` is less than the current number of node groups (shards). A list of shard (node group) IDs to remove from the replication group.
The following procedure describes the process for deleting one or more shards.

**Example - Removing Shards**

The following example removes two node groups from the Redis (cluster mode enabled) cluster `my-cluster`, so there are a total of two node groups when the operation completes. The keyspaces from the removed shards are distributed evenly over the remaining shards.

```plaintext
https://elasticache.us-east-2.amazonaws.com/
  ?Action=ModifyReplicationGroupShardConfiguration
  &ApplyImmediately=true
  &NodeGroupCount=2
  &ReplicationGroupId=my-cluster
  &NodeGroupsToRemove.member.1=0002
  &NodeGroupsToRemove.member.2=0003
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20171102T192317Z
  &X-Amz-Credential=<credential>
```

For more information, see `ModifyReplicationGroupShardConfiguration` in the ElastiCache API Reference.

**Online Shard Rebalancing**

You can rebalance shards in your Redis (cluster mode enabled) cluster using the AWS Management Console, AWS CLI, or ElastiCache API.

**Topics**

- Online Shard Rebalancing (Console) (p. 287)
- Online Shard Rebalancing (AWS CLI) (p. 287)
- Online Shard Rebalancing (ElastiCache API) (p. 288)

**Online Shard Rebalancing (Console)**

The following process describes how to reconfigure the shards in your Redis (cluster mode enabled) cluster by rebalancing shards using the AWS Management Console.

**To rebalance the keyspaces among the shards on your Redis (cluster mode enabled) cluster**

2. From the navigation pane, choose Redis.
3. Choose the name, not the box to the left of the name, of the Redis (cluster mode enabled) cluster that you want to rebalance.

   **Tip**

   Redis (cluster mode enabled) clusters have a value of 1 or greater in the *Shards* column.

5. When prompted, choose Rebalance. You might see a message similar to this one:

   ```plaintext
   Slots in the replication group are uniformly distributed. Nothing to do. (Service: AmazonElastiCache; Status Code: 400; Error Code: InvalidReplicationGroupState; Request ID: 2246cebd-9721-11e7-8d5b-e1b0f086c8cf). If you do, choose Cancel.
   ```

**Online Shard Rebalancing (AWS CLI)**

Use the following parameters with `modify-replication-group-shard-configuration`.

API Version 2015-02-02

287
The following process describes how to reconfigure the shards in your Redis (cluster mode enabled) cluster by rebalancing shards using the AWS CLI.

**Example - Rebalancing the Shards in a Cluster**

The following example rebalances the slots in the Redis (cluster mode enabled) cluster `my-cluster` so that the slots are distributed as equally as possible. The value of `--node-group-count (4)` is the number of shards currently in the cluster.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group-shard-configuration \
  --replication-group-id my-cluster \
  --node-group-count 4 \
  --apply-immediately
```

For Windows:

```bash
aws elasticache modify-replication-group-shard-configuration ^
  --replication-group-id my-cluster ^
  --node-group-count 4 ^
  --apply-immediately
```

For more information, see [modify-replication-group-shard-configuration](https://aws.amazon.com/documentation/elasticache/what-elasticache/) in the AWS CLI documentation.

**Online Shard Rebalancing (ElastiCache API)**

You can use the ElastiCache API to reconfigure the shards in your Redis (cluster mode enabled) cluster online by using the `ModifyReplicationGroupShardConfiguration` operation.

Use the following parameters with `ModifyReplicationGroupShardConfiguration`.

**Parameters**

- **ApplyImmediately=true** – Required. Specifies the shard reconfiguration operation is to be started immediately.
- **ReplicationGroupId** – Required. Specifies which replication group (cluster) the shard reconfiguration operation is to be performed on.
- **NodeGroupCount** – Required. To rebalance the keyspaces across all shards in the cluster, this value must be the same as the current number of shards.

The following process describes how to reconfigure the shards in your Redis (cluster mode enabled) cluster by rebalancing the shards using the ElastiCache API.
Example - Rebalancing a Cluster

The following example rebalances the slots in the Redis (cluster mode enabled) cluster `my-cluster` so that the slots are distributed as equally as possible. The value of `NodeGroupCount` (4) is the number of shards currently in the cluster.

```plaintext
https://elasticache.us-east-2.amazonaws.com/
  ?Action=ModifyReplicationGroupShardConfiguration
  &ApplyImmediately=true
  &NodeGroupCount=4
  &ReplicationGroupId=my-cluster
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20171002T192317Z
  &X-Amz-Credential=<credential>
```

For more information, see `ModifyReplicationGroupShardConfiguration` in the ElastiCache API Reference.

Online Vertical Scaling by Modifying Node Type

By using online vertical scaling with Amazon ElastiCache for Redis version 3.2.10 or newer, you can scale your Redis clusters dynamically with no downtime. This allows your Redis cluster to serve requests even while scaling.

You can do the following:

- **Scale up** – Increase read and write capacity by adjusting the node type of your Redis cluster to use a larger node type.

  ElastiCache dynamically resizes your cluster while remaining online and serving requests.

- **Scale down** – Reduce read and write capacity by adjusting the node type down to use a smaller node. Again, ElastiCache dynamically resizes your cluster while remaining online and serving requests. In this case, you reduce costs by downsizing the node.

**Note**
The scale up and scale down processes rely on creating clusters with newly selected node types and synchronizing the new nodes with the previous ones. To ensure a smooth scale up/down flow, do the following:

- Ensure you have sufficient ENI (Elastic Network Interface) capacity. If scaling down, ensure the smaller node has sufficient memory to absorb expected traffic.

  For best practices on memory management, see `Managing Reserved Memory (p. 466)`.

- While the vertical scaling process is designed to remain fully online, it does rely on synchronizing data between the old node and the new node. We recommend that you initiate scale up/down during hours when you expect data traffic to be at its minimum.

- Test your application behavior during scaling in a staging environment, if possible.

Contents

- Online Scaling Up (p. 290)
  - Scaling Up Redis Cache Clusters (Console) (p. 290)
  - Scaling Up Redis Cache Clusters (AWS CLI) (p. 255)
  - Scaling Up Redis Cache Clusters (ElastiCache API) (p. 292)
- Online Scaling Down (p. 294)
Online Scaling Up

Topics
- Scaling Up Redis Cache Clusters (Console) (p. 290)
- Scaling Up Redis Cache Clusters (AWS CLI) (p. 255)
- Scaling Up Redis Cache Clusters (ElastiCache API) (p. 292)

Scaling Up Redis Cache Clusters (Console)

The following procedure describes how to scale up a Redis cluster using the ElastiCache Management Console. During this process, your Redis cluster will continue to serve requests with minimal downtime.

To scale up a Redis cluster (console)

2. From the navigation pane, choose Redis.
3. From the list of clusters, choose the cluster.
4. Choose Modify.
5. In the Modify Cluster wizard:
   - Choose the node type you want to scale to from the Node type list. To scale up, select a node type larger than your existing node.
6. If you want to perform the scale-up process right away, choose the Apply immediately box. If the Apply immediately box is not chosen, the scale-up process is performed during this cluster's next maintenance window.
7. Choose Modify.

If you chose Apply immediately in the previous step, the cluster's status changes to modifying. When the status changes to available, the modification is complete and you can begin using the new cluster.

Scaling Up Redis Cache Clusters (AWS CLI)

The following procedure describes how to scale up a Redis cache cluster using the AWS CLI. During this process, your Redis cluster will continue to serve requests with minimal downtime.

To scale up a Redis cache cluster (AWS CLI)

1. Determine the node types you can scale up to by running the AWS CLI list-allowed-node-type-modifications command with the following parameter.

   For Linux, macOS, or Unix:

   ```bash
   aws elasticache list-allowed-node-type-modifications \
   --replication-group-id my-replication-group-id
   ```

   For Windows:

   ```bash
   aws elasticache list-allowed-node-type-modifications \n   --replication-group-id my-replication-group-id
   ```

API Version 2015-02-02
290
aws elasticache list-allowed-node-type-modifications ^
--replication-group-id my-replication-group-id

Output from the above command looks something like this (JSON format).

```
{
    "ScaleUpModifications": [ 
        "cache.m3.2xlarge",
        "cache.m3.large",
        "cache.m3.xlarge",
        "cache.m4.10xlarge",
        "cache.m4.2xlarge",
        "cache.m4.4xlarge",
        "cache.m4.large",
        "cache.m4.xlarge",
        "cache.r3.2xlarge",
        "cache.r3.4xlarge",
        "cache.r3.8xlarge",
        "cache.r3.large",
        "cache.r3.xlarge"
    ],
    "ScaleDownModifications": [ 
        "cache.t2.micro",
        "cache.t2.small",
        "cache.t2.medium",
        "cache.t1.small"
    ],
}
```

For more information, see list-allowed-node-type-modifications in the AWS CLI Reference.

2. Modify your replication group to scale up to the new, larger node type, using the AWS CLI modify-replication-group command and the following parameters.

- **--replication-group-id** – The name of the replication group you are scaling up to.
- **--cache-node-type** – The new node type you want to scale the cache cluster. This value must be one of the node types returned by the list-allowed-node-type-modifications command in step 1.
- **--cache-parameter-group-name** – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
- **--apply-immediately** – Causes the scale-up process to be applied immediately. To postpone the scale-up process to the cluster’s next maintenance window, use the --no-apply-immediately parameter.

For Linux, macOS, or Unix:

```
aws elasticache modify-replication-group  \
--replication-group-id my-redis-cluster  \ 
--cache-node-type cache.m3.xlarge  \
--apply-immediately
```

For Windows:

```powershell
aws elasticache modify-replication-group ^
--replication-group-id my-redis-cluster ^
```
Output from the above command looks something like this (JSON format).

```
{
  "ReplicationGroup": {
    "Status": "modifying",
    "Description": "my-redis-cluster",
    "NodeGroups": [
      {
        "Status": "modifying",
        "Slots": "0-16383",
        "NodeGroupId": "0001",
        "NodeGroupMembers": [
          {
            "PreferredAvailabilityZone": "us-east-1f",
            "CacheNodeId": "0001",
            "CacheClusterId": "my-redis-cluster-0001-001"
          },
          {
            "PreferredAvailabilityZone": "us-east-1d",
            "CacheNodeId": "0001",
            "CacheClusterId": "my-redis-cluster-0001-002"
          }
        ]
      }
    ],
    "ConfigurationEndpoint": {
      "Port": 6379,
      "Address": "my-redis-cluster.r7gdfi.clusterconfig.use1.cache.amazonaws.com"
    },
    "ClusterEnabled": true,
    "ReplicationGroupId": "my-redis-cluster",
    "SnapshotRetentionLimit": 1,
    "AutomaticFailover": "enabled",
    "SnapshotWindow": "07:30-08:30",
    "MemberClusters": [
      "my-redis-cluster-0001-001",
      "my-redis-cluster-0001-002"
    ],
    "CacheNodeType": "cache.m3.xlarge",
    "PendingModifiedValues": {}
  }
}
```

For more information, see modify-replication-group in the AWS CLI Reference.

3. If you used the --apply-immediately, check the status of the cache cluster using the AWS CLI describe-cache-clusters command with the following parameter. When the status changes to available, you can begin using the new, larger cache cluster node.

**Scaling Up Redis Cache Clusters (ElastiCache API)**

The following process scales your cache cluster from its current node type to a new, larger node type using the ElastiCache API. During this process, until the status changes from modifying to available, all reads and writes between your application and the primary cache cluster are blocked. However, reads from the read replica cache clusters continue uninterrupted.

The amount of time it takes to scale up to a larger node type varies, depending upon your node type and the amount of data in your current cache cluster.
To scale up a Redis Cache Cluster (ElastiCache API)

1. Determine which node types you can scale up to using the ElastiCache API ListAllowedNodeTypeModifications action with the following parameter.

   - ReplicationGroupId – the name of the replication group. Use this parameter to describe a specific replication group rather than all replication groups.

   ```
   https://elasticache.us-west-2.amazonaws.com/
   ?Action/ListAllowedNodeTypeModifications
   &ReplicationGroupId=MyReplGroup
   &Version=2015-02-02
   &SignatureVersion=4
   &SignatureMethod=HmacSHA256
   &Timestamp=20150202T192317Z
   &X-Amz-Credential=<credential>
   
   For more information, see ListAllowedNodeTypeModifications in the Amazon ElastiCache API Reference.
   ```

2. Scale your current replication group up to the new node type using the ModifyReplicationGroup ElastiCache API action and with the following parameters.

   - ReplicationGroupId – the name of the replication group.
   - CacheNodeType – the new, larger node type of the cache clusters in this replication group. This value must be one of the instance types returned by the ListAllowedNodeTypeModifications action in step 1.
   - CacheParameterGroupName – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
   - ApplyImmediately – Set to true to causes the scale-up process to be applied immediately. To postpone the scale-up process to the next maintenance window, use ApplyImmediately=false.

   ```
   https://elasticache.us-west-2.amazonaws.com/
   ?Action=ModifyReplicationGroup
   &ApplyImmediately=true
   &CacheNodeType=cache.m3.2xlarge
   &CacheParameterGroupName=redis32-m3-2xl
   &ReplicationGroupId=myReplGroup
   &SignatureVersion=4
   &SignatureMethod=HmacSHA256
   &Timestamp=20141201T220302Z
   &Version=2014-12-01
   &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 ModifyReplicationGroup in the Amazon ElastiCache API Reference.
   ```

3. If you used ApplyImmediately=true, monitor the status of the replication group using the ElastiCache API DescribeReplicationGroups action with the following parameters. When the status changes from modifying to available, you can begin writing to your new, scaled up replication group.

   ```
   API Version 2015-02-02
   293```
Scaling Clusters in Redis (Cluster Mode Enabled)

- ReplicationGroupId - the name of the replication group. Use this parameter to describe a particular replication group rather than all replication groups.

https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeReplicationGroups
&ReplicationGroupId=MyRepGroup
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>

For more information, see DescribeReplicationGroups in the Amazon ElastiCache API Reference.

Online Scaling Down

Topics
- Scaling Down Redis Cache Clusters (Console) (p. 294)
- Scaling Down Redis Cache Clusters (AWS CLI) (p. 294)
- Scaling Down Redis Cache Clusters (ElastiCache API) (p. 297)

Scaling Down Redis Cache Clusters (Console)

The following procedure describes how to scale down a Redis cluster using the ElastiCache Management Console. During this process, your Redis cluster will continue to serve requests with minimal downtime.

To scale down a Redis cluster (console)

2. From the navigation pane, choose Redis.
3. From the list of clusters, choose your preferred cluster.
4. Choose Modify.
5. In the Modify Cluster wizard:
   - Choose the node type you want to scale to from the Node type list. To scale down, select a node type smaller than your existing node. Note that not all node types are available to scale down to.
6. If you want to perform the scale down process right away, choose the Apply immediately box. If the Apply immediately box is not chosen, the scale-down process is performed during this cluster’s next maintenance window.
7. Choose Modify.

   If you chose Apply immediately in the previous step, the cluster’s status changes to modifying. When the status changes to available, the modification is complete and you can begin using the new cluster.

Scaling Down Redis Cache Clusters (AWS CLI)

The following procedure describes how to scale down a Redis cache cluster using the AWS CLI. During this process, your Redis cluster will continue to serve requests with minimal downtime.
To scale down a Redis cache cluster (AWS CLI)

1. Determine the node types you can scale down to by running the AWS CLI list-allowed-node-type-modifications command with the following parameter.

   For Linux, macOS, or Unix:

   ```bash
   aws elasticache list-allowed-node-type-modifications \
     --replication-group-id my-replication-group-id
   ```

   For Windows:

   ```bash
   aws elasticache list-allowed-node-type-modifications ^
     --replication-group-id my-replication-group-id
   ```

   Output from the above command looks something like this (JSON format).

   ```json
   {
     "ScaleUpModifications": [
       "cache.m3.2xlarge",
       "cache.m3.large",
       "cache.m3.xlarge",
       "cache.m4.10xlarge",
       "cache.m4.2xlarge",
       "cache.m4.4xlarge",
       "cache.m4.large",
       "cache.m4.xlarge",
       "cache.r3.2xlarge",
       "cache.r3.4xlarge",
       "cache.r3.8xlarge",
       "cache.r3.large",
       "cache.r3.xlarge"
     ],
     "ScaleDownModifications": [
       "cache.t2.micro",
       "cache.t2.small",
       "cache.t2.medium",
       "cache.t1.small"
     ]
   }
   ```

   For more information, see list-allowed-node-type-modifications in the AWS CLI Reference.

2. Modify your replication group to scale down to the new, smaller node type, using the AWS CLI modify-replication-group command and the following parameters.

   • `--replication-group-id` – The name of the replication group you are scaling down to.
   • `--cache-node-type` – The new node type you want to scale the cache cluster. This value must be one of the node types returned by the list-allowed-node-type-modifications command in step 1.
   • `--cache-parameter-group-name` – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
   • `--apply-immediately` – Causes the scale-up process to be applied immediately. To postpone the scale-down process to the cluster's next maintenance window, use the `--no-apply-immediately` parameter.
For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group \
  --replication-group-id my-redis-cluster \
  --cache-node-type cache.t2.micro \
  --apply-immediately
```

For Windows:

```bash
aws elasticache modify-replication-group ^
  --replication-group-id my-redis-cluster ^
  --cache-node-type cache.t2.micro ^
  --apply-immediately
```

Output from the above command looks something like this (JSON format).

```json
{
  "ReplicationGroup": {
    "Status": "modifying",
    "Description": "my-redis-cluster",
    "NodeGroups": [
      {
        "Status": "modifying",
        "Slots": "0-16383",
        "NodeGroupId": "0001",
        "NodeGroupMembers": [
          {
            "PreferredAvailabilityZone": "us-east-1f",
            "CacheNodeId": "0001",
            "CacheClusterId": "my-redis-cluster-0001-001"
          },
          {
            "PreferredAvailabilityZone": "us-east-1d",
            "CacheNodeId": "0001",
            "CacheClusterId": "my-redis-cluster-0001-002"
          }
        ]
      }
    ],
    "ConfigurationEndpoint": {
      "Port": 6379,
      "Address": "my-redis-cluster.r7gdfi.clustercfg.use1.cache.amazonaws.com"
    },
    "ClusterEnabled": true,
    "ReplicationGroupId": "my-redis-cluster",
    "SnapshotRetentionLimit": 1,
    "AutomaticFailover": "enabled",
    "SnapshotWindow": "07:30-08:30",
    "MemberClusters": [
      "my-redis-cluster-0001-001",
      "my-redis-cluster-0001-002"
    ],
    "CacheNodeType": "cache.t2.micro",
    "PendingModifiedValues": {}
  }
}
```

For more information, see `modify-replication-group` in the AWS CLI Reference.
3. If you used the --apply-immediately, check the status of the cache cluster using the AWS CLI describe-cache-clusters command with the following parameter. When the status changes to available, you can begin using the new, smaller cache cluster node.

### Scaling Down Redis Cache Clusters (ElastiCache API)

The following process scales your replication group from its current node type to a new, smaller node type using the ElastiCache API. During this process, your Redis cluster will continue to serve requests with minimal downtime.

The amount of time it takes to scale down to a smaller node type varies, depending upon your node type and the amount of data in your current cache cluster.

#### Scaling Down (ElastiCache API)

1. Determine which node types you can scale down to using the ElastiCache API ListAllowedNodeTypeModifications action with the following parameter.

   • ReplicationGroupId – the name of the replication group. Use this parameter to describe a specific replication group rather than all replication groups.

   ```
   https://elasticache.us-west-2.amazonaws.com/
   ?Action=ListAllowedNodeTypeModifications
   &ReplicationGroupId=MyReplGroup
   &Version=2015-02-02
   &SignatureVersion=4
   &SignatureMethod=HmacSHA256
   &Timestamp=20150202T192317Z
   &X-Amz-Credential=<credential>
   
   For more information, see ListAllowedNodeTypeModifications in the Amazon ElastiCache API Reference.
   ```

2. Scale your current replication group down to the new node type using the ModifyReplicationGroup ElastiCache API action and with the following parameters.

   • ReplicationGroupId – the name of the replication group.
   • CacheNodeType – the new, smaller node type of the cache clusters in this replication group. This value must be one of the instance types returned by the ListAllowedNodeTypeModifications action in step 1.
   • CacheParameterGroupName – [Optional] Use this parameter if you are using reserved-memory to manage your cluster's reserved memory. Specify a custom cache parameter group that reserves the correct amount of memory for your new node type. If you are using reserved-memory-percent you can omit this parameter.
   • ApplyImmediately – Set to true to causes the scale-down process to be applied immediately. To postpone the scale-down process to the next maintenance window, use ApplyImmediately=false.

   ```
   https://elasticache.us-west-2.amazonaws.com/
   ?Action=ModifyReplicationGroup
   &ApplyImmediately=true
   &CacheNodeType=cache.t2.micro
   &CacheParameterGroupName=redis32-m3-2x1
   &ReplicationGroupId=myReplGroup
   &SignatureVersion=4
   &SignatureMethod=HmacSHA256
   &Timestamp=20141201T220302Z
   ```
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 Redis Specific Parameters (p. 315).

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 Redis Node-Type Specific Parameters (p. 336).

**Note**
For a list of Memcached-specific parameters, see Memcached Specific Parameters.

**Topics**
- Parameter Management (p. 299)
- Cache Parameter Group Tiers (p. 300)
- Creating a Parameter Group (p. 300)
- Listing Parameter Groups by Name (p. 304)
- Listing a Parameter Group's Values (p. 309)
- Modifying a Parameter Group (p. 310)
- Deleting a Parameter Group (p. 313)
- Redis Specific Parameters (p. 315)
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 Redis version 3.2.10, you can only use parameter groups, default or custom, from the Redis3.2 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 Redis Specific Parameters (p. 315). For information on rebooting a cluster, see Rebooting a Cluster (p. 103).
- You can associate parameter groups with Redis global datastores. Global datastores are a collection of one or more clusters that span AWS Regions. In this case, the parameter group is shared by all clusters that make up the global datastore. Any modifications to the parameter group of the primary cluster are replicated to all remaining clusters in the global datastore. For more information, see Replication Across AWS Regions Using Global Datastore (p. 130).

You can check if a parameter group is part of a global datastore by looking in these locations:
- On the ElastiCache console on the Parameter Groups page, the yes/no Global attribute
- The yes/no IsGlobal property of the CacheParameterGroup API operation
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 Redis Specific Parameters (p. 315).

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 redis3.2, 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. 310).

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.
• --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 myRed28 using the redis2.8 family as the template.

For Linux, macOS, or Unix:

```bash
aws elasticache create-cache-parameter-group \
  --cache-parameter-group-name myRed28 \
  --cache-parameter-group-family redis2.8 \
  --description "My first parameter group"
```

For Windows:

```bash
aws elasticache create-cache-parameter-group ^  \
  --cache-parameter-group-name myRed28 ^  \
  --cache-parameter-group-family redis2.8 ^  \
  --description "My first parameter group"
```

The output from this command should look something like this.

```json
{
  "CacheParameterGroup": {
    "CacheParameterGroupName": "myRed28",
    "CacheParameterGroupFamily": "redis2.8",
    "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. 310).

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, redis2.8.
• Description — A user supplied description for the parameter group.
Example

The following example creates a parameter group named `myRed28` using the redis2.8 family as the template.

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateCacheParameterGroup
  &CacheParameterGroupFamily=redis2.8
  &CacheParameterGroupName=myRed28
  &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>myRed28</CacheParameterGroupName>
      <CacheParameterGroupFamily>redis2.8</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. 310).

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 myRed28.

For Linux, macOS, or Unix:

```bash
aws elasticache describe-cache-parameter-groups
   --cache-parameter-group-name myRed28
```

For Windows:

```bash
aws elasticache describe-cache-parameter-groups ^
   --cache-parameter-group-name myRed28
```

The output of this command will look something like this, listing the name, family, and description for the parameter group.

```
{
   "CacheParameterGroups": [
   {
      "CacheParameterGroupName": "myRed28",
      "CacheParameterGroupFamily": "redis2.8",
      "Description": "My first parameter group"
   }
   ]
}
```

Example

The following sample code lists the parameter group myRed56 for parameter groups running on Redis engine version 5.0.6 onwards. If the parameter group is part of a Replication Across AWS Regions Using Global Datastore (p. 130), the IsGlobal property value returned in the output will be Yes.

For Linux, macOS, or Unix:
aws elasticache describe-cache-parameter-groups \
   --cache-parameter-group-name myRed56

For Windows:

aws elasticache describe-cache-parameter-groups ^ \
   --cache-parameter-group-name myRed56

The output of this command will look something like this, listing the name, family, isGlobal and description for the parameter group.

```json
{
   "CacheParameterGroups": [
   {
      "CacheParameterGroupName": "myRed56",
      "CacheParameterGroupFamily": "redis5.0",
      "Description": "My first parameter group",
      "IsGlobal": "yes"
   }
   ]
}
```

**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 Datastore (isGlobal), for each parameter group.

```json
{
   "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"
   }
   ]
}
```
Listing Parameter Groups by Name

For more information, see `describe-cache-parameter-groups`.

**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 up to 10 parameter groups.

```
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.

```
<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 5.6 parameter group</Description>
</CacheParameterGroup>
</CacheParameterGroups>
</DescribeCacheParameterGroupsResult>
</DescribeCacheParameterGroupsResponse>
```
Example

The following sample code lists the parameter group *myRed28*.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheParameterGroups
&CacheParameterGroupName=myRed28
&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, and description.

```
  <DescribeCacheParameterGroupsResult>
    <CacheParameterGroups>
      <CacheParameterGroup>
        <CacheParameterGroupName>myRed28</CacheParameterGroupName>
        <CacheParameterGroupFamily>redis2.8</CacheParameterGroupFamily>
        <Description>My custom Redis 2.8 parameter group</Description>
      </CacheParameterGroup>
    </CacheParameterGroups>
  </DescribeCacheParameterGroupsResult>
  <ResponseMetadata>
    <RequestId>3540cc3d-af48-11e0-97f9-279771c4477e</RequestId>
  </ResponseMetadata>
</DescribeCacheParameterGroupsResponse>
```

Example

The following sample code lists the parameter group *myRed56*.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheParameterGroups
&CacheParameterGroupName=myRed56
&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 whether the parameter group is part of a Global Datastore (isGlobal).

```
  <DescribeCacheParameterGroupsResult>
    <CacheParameterGroups>
    </CacheParameterGroups>
  </DescribeCacheParameterGroupsResult>
  <ResponseMetadata>
    <RequestId>3540cc3d-af48-11e0-97f9-279771c4477e</RequestId>
  </ResponseMetadata>
</DescribeCacheParameterGroupsResponse>
```
Example

The following sample code lists up to 10 parameter groups.

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, whether the parameter group is part of a Global Datastore (isGlobal), for each parameter group.

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 myRedis28.

For Linux, macOS, or Unix:

```bash
aws elasticache describe-cache-parameters \
  --cache-parameter-group-name myRedis28
```

For Windows:

```bash
aws elasticache describe-cache-parameters ^
  --cache-parameter-group-name myRedis28
```

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 myRed28.

```bash
https://elasticache.us-west-2.amazonaws.com/
```
The response from this action will look something like this. This response has been truncated.

```
  <DescribeCacheParametersResult>
    <CacheClusterClassSpecificParameters>
      <CacheNodeTypeSpecificParameter>
        <DataType>integer</DataType>
        <Source>system</Source>
        <IsModifiable>false</IsModifiable>
        <Description>The maximum configurable amount of memory to use to store items, in megabytes.</Description>
        <CacheNodeTypeSpecificValues>
          <CacheNodeTypeSpecificValue>
            <Value>1000</Value>
            <CacheClusterClass>cache.c1.medium</CacheClusterClass>
          </CacheNodeTypeSpecificValue>
          <CacheNodeTypeSpecificValue>
            <Value>6000</Value>
            <CacheClusterClass>cache.c1.xlarge</CacheClusterClass>
          </CacheNodeTypeSpecificValue>
          <CacheNodeTypeSpecificValue>
            <Value>7100</Value>
            <CacheClusterClass>cache.m1.large</CacheClusterClass>
          </CacheNodeTypeSpecificValue>
          <CacheNodeTypeSpecificValue>
            <Value>1300</Value>
            <CacheClusterClass>cache.m1.small</CacheClusterClass>
          </CacheNodeTypeSpecificValue>
        </CacheNodeTypeSpecificValues>
      </CacheNodeTypeSpecificParameter>
    </CacheClusterClassSpecificParameters>
  </DescribeCacheParametersResult>
  <ResponseMetadata>
    <RequestId>6d355589-af49-11e0-97f9-279771c4477e</RequestId>
  </ResponseMetadata>
</DescribeCacheParametersResponse>
```

For more information, see [DescribeCacheParameters](http://elasticache.amazonaws.com/doc/2013-06-15/).

## 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 [Redis Specific Parameters](p. 315).

## 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 Redis Specific Parameters (p. 315). 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. 103).

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 Redis Specific Parameters (p. 315)

The following sample code sets the value of two parameters, reserved-memory-percent and cluster-enabled on the parameter group myredis32-on-30. We set reserved-memory-percent to 30 (30 percent) and cluster-enabled to yes so that the parameter group can be used with Redis (cluster mode enabled) clusters (replication groups).

For Linux, macOS, or Unix:

```
aws elasticache modify-cache-parameter-group \
  --cache-parameter-group-name myredis32-on-30 \
  --parameter-name-values \n  ParameterName=reserved-memory-percent,ParameterValue=30 \
  ParameterName=cluster-enabled,ParameterValue=yes
```

For Windows:

```
aws elasticache modify-cache-parameter-group ^
  --cache-parameter-group-name myredis32-on-30 ^
  --parameter-name-values ^
  ParameterName=reserved-memory-percent,ParameterValue=30 ^
  ParameterName=cluster-enabled,ParameterValue=yes
```

Output from this command will look something like this.

```json
{
  "CacheParameterGroupName": "my-redis32-on-30"
}
```
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. 103).

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 Redis Specific Parameters (p. 315)

The following sample code sets the value of two parameters, reserved-memory-percent and cluster-enabled on the parameter group myredis32-on-30. We set reserved-memory-percent to 30 (30 percent) and cluster-enabled to yes so that the parameter group can be used with Redis (cluster mode enabled) clusters (replication groups).

https://elasticache.us-west-2.amazonaws.com/
?Action=ModifyCacheParameterGroup
&CacheParameterGroupName=myredis32-on-30
&ParameterNameValues.member.1.ParameterName=reserved-memory-percent
&ParameterNameValues.member.1.ParameterValue=30
&ParameterNameValues.member.2.ParameterName=cluster-enabled
&ParameterNameValues.member.2.ParameterValue=yes
&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. 103).
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 myRed28
```

For Windows:

```
aws elasticache delete-cache-parameter-group ^
  --cache-parameter-group-name myRed28
```

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 `myRed28` parameter group.

```xml
https://elasticache.us-west-2.amazonaws.com/
?Action=DeleteCacheParameterGroup
&CacheParameterGroupName=myRed28
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&Version=2015-02-02
&X-Amz-Credential=<credential>

For more information, see `DeleteCacheParameterGroup`.```
Redis Specific Parameters

If you do not specify a parameter group for your Redis cluster, then a default parameter group appropriate to your engine version will be used. You can't change the values of any parameters in the default parameter group. However, you can create a custom parameter group and assign it to your cluster at any time as long as the values of conditionally modifiable parameters are the same in both parameter groups. For more information, see Creating a Parameter Group (p. 300).

Topics

- Redis 5.0.3 Parameter Changes (p. 316)
- Redis 5.0.0 Parameter Changes (p. 317)
- Redis 4.0.10 Parameter Changes (p. 320)
- Redis 3.2.10 Parameter Changes (p. 323)
- Redis 3.2.6 Parameter Changes (p. 323)
- Redis 3.2.4 Parameter Changes (p. 323)
- Redis 2.8.24 (Enhanced) Added Parameters (p. 326)
- Redis 2.8.23 (Enhanced) Added Parameters (p. 326)
- Redis 2.8.22 (Enhanced) Added Parameters (p. 328)
- Redis 2.8.21 Added Parameters (p. 328)
- Redis 2.8.19 Added Parameters (p. 328)
- Redis 2.8.6 Added Parameters (p. 329)
- Redis 2.6.13 Parameters (p. 331)
- Redis Node-Type Specific Parameters (p. 336)

Note
Because the newer Redis versions provide a better and more stable user experience, Redis versions 2.6.13, 2.8.6, and 2.8.19 are deprecated when using the ElastiCache console. We recommend against using these Redis versions. If you need to use one of them, work with the AWS CLI or ElastiCache API.
For more information, see the following topics:

<table>
<thead>
<tr>
<th></th>
<th>AWS CLI</th>
<th>ElastiCache API</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Cluster</td>
<td>Creating a Cluster (AWS CLI) (p. 87)</td>
<td>Creating a Cluster (ElastiCache API) (p. 88)</td>
</tr>
<tr>
<td></td>
<td>You can't use this action to create a replication group with cluster mode enabled.</td>
<td>You can't use this action to create a replication group with cluster mode enabled.</td>
</tr>
<tr>
<td>Modify Cluster</td>
<td>Using the AWS CLI (p. 101)</td>
<td>Using the ElastiCache API (p. 102)</td>
</tr>
<tr>
<td></td>
<td>You can't use this action to create a replication group with cluster mode enabled.</td>
<td>You can't use this action to create a replication group with cluster mode enabled.</td>
</tr>
<tr>
<td>Create Replication Group</td>
<td>Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch (AWS CLI) (p. 169)</td>
<td>Creating a Redis (cluster mode disabled) Replication Group from Scratch (ElastiCache API) (p. 172)</td>
</tr>
</tbody>
</table>
### Redis 5.0.3 Parameter Changes

**Parameter group family:** redis5.0

Redis 5.0 default parameter groups

- `default.redis5.0` – Use this parameter group, or one derived from it, for Redis (cluster mode disabled) clusters and replication groups.
- `default.redis5.0.cluster.on` – Use this parameter group, or one derived from it, for Redis (cluster mode enabled) clusters and replication groups.

**Parameters added in Redis 5.0.3**

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>rename-commands</td>
<td>Default: none</td>
<td>A space-separated list of renamed Redis commands. The following is a restricted list of</td>
</tr>
<tr>
<td></td>
<td>Type: string</td>
<td>commands available for renaming:</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td>APPEND AUTH BITCOUNT BITFIELD BITPOS BLPOP BRPOP BRPOPLPUSH BZPOPMIN BZPOPMAX CLIENT CLUSTER</td>
</tr>
<tr>
<td></td>
<td>Changes take effect:</td>
<td>COMMAND DBSIZE DECR DECRBY DEL DISCARD DUMP ECHO EVAL EVALSHA EXEC EXISTS EXPIRE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXPIREAT FLUSHALL FLUSHDB GEOADD GEOHASH GEOPOS GEODIST GEORADIUS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GEORADIUSBYMEMBER GET GETBIT GETRANGE GETSET HDEL HEXISTS HGET HGETALL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HINCRBY HINCRBYFLOAT HKEYS HLEN HMGET HSET HSETNX HSTRLEN HVALS INCR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INCRBY INCRBYFLOAT INFO KEYS LASTSAVE LINDEX LINSERT LLEN LPUSH LPUSHX LRENAME LREM LSET</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LTRIM MEMORY MGET MONITOR MOVE MSET MSETNX MULTI OBJECT PERSIST PEXPIRE PEXPIREAT PFADD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PFCOUNT PFMERGE PING PSETEX PSUBSCRIBE PUBSUB PTTL PUBLISH PUNSUBSCRIBE RANDOMKEY READONLY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>READWRITE RENAME RENAMEEX RENAMEEX RENAMEINDEX RESTORE ROLE RPOP RPOPLPUSH RPOPLPUSH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RPUSH RPUSHX SADD SCARD SCRIPT SDIFF SDIFFSTORE SELECT SET SETBIT SETEX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SETNX SETRANGE SINTER SINTERSTORE SISMEMBER SLOWLOG SMEMBERS SMOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SORT SPOP SRANDMEMBER SREM STRENSN SUBSCRIBE SUNION SUNIONSTORE SWAPDB TIME TOUCH TTL TYPE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UNSUBSCRIBE</td>
</tr>
</tbody>
</table>
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNLINK</td>
<td>UNWATCH WAIT WATCH ZADD ZCARD</td>
<td>ZCOUNT ZINCRBY ZINTERSTORE ZLEXCOUNT ZPOPMAX ZPOPMIN ZRANGE ZRANGEBYLEX ZREVRANGEBYLEX ZREVRANGEBYSCORE ZRANK ZREM ZREMRANGE ZREMRANGEBYLEX ZREMRANGEBYSCORE ZREVRANGE ZREVRANGEBYSCORE ZREVRANK ZSCORE ZUNIONSTORE SCAN SSCAN HSCAN ZSCAN XINFO XADD XTRIM XDEL XRANGE XREVRANGE XLEN XREAD XGROUP XREADGROUP XACK XCLAIM XPENDING GEORADIUS_RO GEORADIUSBYMEMBER_RO LOLWUT XSETID SUBSTR</td>
</tr>
</tbody>
</table>

For more information, see [ElastiCache for Redis Version 5.0.3 (Enhanced)](p. 47).

### Redis 5.0.0 Parameter Changes

**Parameter group family:** redis5.0

Redis 5.0 default parameter groups

- `default.redis5.0` – Use this parameter group, or one derived from it, for Redis (cluster mode disabled) clusters and replication groups.
- `default.redis5.0.cluster.on` – Use this parameter group, or one derived from it, for Redis (cluster mode enabled) clusters and replication groups.

**Parameters added in Redis 5.0**

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>stream-node-max-bytes</code></td>
<td>Permitted values: 0+ Default: 4096</td>
<td>Type: integer Modifiable: Yes Changes take effect: Immediately The stream data structure is a radix tree of nodes that encode multiple items inside. Use this configuration to specify the maximum size of a single node in radix tree in Bytes. If set to 0, the size of the tree node is unlimited.</td>
</tr>
<tr>
<td><code>stream-node-max-entries</code></td>
<td>Permitted values: 0+ Default: 100</td>
<td>Type: integer Modifiable: Yes Changes take effect: Immediately The stream data structure is a radix tree of nodes that encode multiple items inside. Use this configuration to specify the maximum number of items a single node can contain before switching to a new node when appending new stream entries. If set to 0, the number of items in the tree node is unlimited.</td>
</tr>
<tr>
<td><code>active-defrag-max-scan-fields</code></td>
<td>Permitted values: 1 to 1000000 Default: 1000</td>
<td>Maximum number of set/hash/zset/list fields that will be processed from the main dictionary scan.</td>
</tr>
</tbody>
</table>
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<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:</td>
<td>Immediately</td>
</tr>
<tr>
<td>lua-replicate-commands</td>
<td>Permitted values: yes/no</td>
<td>Always enable Lua effect replication or not in Lua scripts</td>
</tr>
<tr>
<td></td>
<td>Default: yes</td>
<td></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:</td>
<td>Immediately</td>
</tr>
<tr>
<td>replica-ignore-maxmemory</td>
<td>Default: yes</td>
<td>Determines if replica ignores maxmemory setting by not evicting items</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>by not evicting items independent from the master</td>
</tr>
</tbody>
</table>

Redis has renamed several parameters in engine version 5.0 in response to community feedback. For more information, see What’s New in Redis 5?. The following table lists the new names and how they map to previous versions.

### Parameters renamed in Redis 5.0

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>replica-lazy-flush</td>
<td>Default: no</td>
<td>Performs an asynchronous flushDB during replica sync.</td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Former name: slave-lazy-flush</td>
<td></td>
</tr>
<tr>
<td>client-output-buffer-limit</td>
<td>Default: For values see Redis Node-Type Specific Parameters (p. 336)</td>
<td>For Redis read replicas: If a client's output buffer reaches the specified number of bytes, the client will be disconnected.</td>
</tr>
<tr>
<td>replica-hard-limit</td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Former name: client-output-buffer-limit-slave-hard-limit</td>
<td></td>
</tr>
<tr>
<td>client-output-buffer-limit</td>
<td>Default: For values see Redis Node-Type Specific Parameters (p. 336)</td>
<td>For Redis read replicas: If a client's output buffer reaches the specified number of bytes, the client will be disconnected, but only if this condition persists for client-output-buffer-limit-replica-soft-seconds.</td>
</tr>
<tr>
<td>replica-soft-limit</td>
<td>Type: integer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Details</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| client-output-buffer-limit-slave-soft-seconds | Default: 60  
Type: integer  
Modifiable: No  
Former name: client-output-buffer-limit-slave-soft-seconds | For Redis read replicas: If a client's output buffer remains at client-output-buffer-limit-replica-soft-limit bytes for longer than this number of seconds, the client will be disconnected.                        |
| replica-allow-chaining                    | Default: no  
Type: string  
Modifiable: No  
Former name: slave-allow-chaining | Determines whether a read replica in Redis can have read replicas of its own.                                                                                                                                |
| min-replicas-to-write                    | Default: 0  
Type: integer  
Modifiable: Yes  
Former name: min-slaves-to-write  
Changes Take Effect: Immediately | The minimum number of read replicas which must be available in order for the primary node to accept writes from clients. If the number of available replicas falls below this number, then the primary node will no longer accept write requests.  
If either this parameter or min-replicas-max-lag is 0, then the primary node will always accept write requests, even if no replicas are available. |
| min-replicas-max-lag                    | Default: 10  
Type: integer  
Modifiable: Yes  
Former name: min-slaves-max-lag  
Changes Take Effect: Immediately | The number of seconds within which the primary node must receive a ping request from a read replica. If this amount of time passes and the primary does not receive a ping, then the replica is no longer considered available.  
If the number of available replicas drops below min-replicas-to-write, then the primary will stop accepting writes at that point.  
If either this parameter or min-replicas-to-write is 0, then the primary node will always accept write requests, even if no replicas are available. |
| close-on-replica-write                    | Default: yes  
Type: boolean  
Modifiable: Yes  
Former name: close-on-slave-write  
Changes Take Effect: Immediately | If enabled, clients who attempt to write to a read-only replica will be disconnected.                                                                                                                         |
Parameters removed in Redis 5.0

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>repl-timeout</td>
<td>Default: 60</td>
<td>Parameter is not available in this version.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
</tbody>
</table>

Redis 4.0.10 Parameter Changes

Parameter group family: redis4.0

Redis 4.0.x default parameter groups

- default.redis4.0 – Use this parameter group, or one derived from it, for Redis (cluster mode disabled) clusters and replication groups.
- default.redis4.0.cluster.on – Use this parameter group, or one derived from it, for Redis (cluster mode enabled) clusters and replication groups.

Parameters changed in Redis 4.0.10

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxmemory-policy</td>
<td>Permitted values: allkeys-lru, volatile-lru, allkeys-lfu, volatile-lfu, allkeys-random, volatile-random, volatile-ttl, noeviction</td>
<td>maxmemory-policy was added in version 2.6.13. In version 4.0.10 two new permitted values are added: allkeys-lfu, which will evict any key using approximated LFU, and volatile-lfu, which will evict using approximated LFU among the keys with an expire set.</td>
</tr>
<tr>
<td></td>
<td>Default: volatile-lru</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Type: string</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes take place: immediately</td>
<td></td>
</tr>
</tbody>
</table>

Parameters Added in Redis 4.0.10

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lazyfree-lazy-eviction</td>
<td>Permitted values: yes/no</td>
<td>Performs an asynchronous delete on evictions.</td>
</tr>
<tr>
<td></td>
<td>Default: no</td>
<td></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 place: immediately</td>
<td></td>
</tr>
<tr>
<td>lazyfree-lazy-expire</td>
<td>Permitted values: yes/no</td>
<td>Performs an asynchronous delete on expired keys.</td>
</tr>
<tr>
<td></td>
<td>Default: no</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Details</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>lazyfree</td>
<td>Permitted values: yes/no</td>
<td>Performs an asynchronous delete for commands which update values.</td>
</tr>
<tr>
<td>lazyserver</td>
<td>Default: no</td>
<td></td>
</tr>
<tr>
<td>lazy-del</td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes take place: immediately</td>
<td></td>
</tr>
<tr>
<td>slave</td>
<td>Permitted values: N/A</td>
<td>Performs an asynchronous flushDB during slave sync.</td>
</tr>
<tr>
<td>lazy-flush</td>
<td>Default: no</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Type: boolean</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Changes take place: N/A</td>
<td></td>
</tr>
</tbody>
</table>

**LFU parameters**

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lfu-log-factor</td>
<td>Permitted values: any integer &gt; 0</td>
<td>Set the log factor, which determines the number of key hits to saturate the key counter.</td>
</tr>
<tr>
<td></td>
<td>Default: 10</td>
<td></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 place: immediately</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lfu-decay-time</td>
<td>Permitted values: any integer</td>
<td>The amount of time in minutes to decrement the key counter.</td>
</tr>
<tr>
<td></td>
<td>Default: 1</td>
<td></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 place: immediately</td>
<td></td>
</tr>
</tbody>
</table>

**Active defragmentation parameters**

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>active-defrag</td>
<td>Permitted values: yes/no</td>
<td>Enabled active defragmentation.</td>
</tr>
<tr>
<td></td>
<td>Default: no</td>
<td></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 place: immediately</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Details</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| active-defrag-ignore-bytes                | Permitted values: 10485760-104857600
Default: 104857600
Type: integer
Modifiable: Yes
Changes take place: immediately         | Minimum amount of fragmentation waste to start active defrag.            |
| active-defrag-threshold-lower             | Permitted values: 1-100
Default: 10
Type: integer
Modifiable: Yes
Changes take place: immediately         | Minimum percentage of fragmentation to start active defrag.              |
| active-defrag-threshold-upper             | Permitted values: 1-100
Default: 100
Type: integer
Modifiable: Yes
Changes take place: immediately         | Maximum percentage of fragmentation at which we use maximum effort.      |
| active-defrag-cycle-min                   | Permitted values: 1-75
Default: 25
Type: integer
Modifiable: Yes
Changes take place: immediately         | Minimal effort for defrag in CPU percentage.                            |
| active-defrag-cycle-max                   | Permitted values: 1-75
Default: 75
Type: integer
Modifiable: Yes
Changes take place: immediately         | Maximal effort for defrag in CPU percentage.                            |

**Client output buffer parameters**
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Default: 1073741824</td>
<td></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 place: immediately</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Permitted values: 1048576-536870912</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Default: 536870912</td>
<td></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 place: immediately</td>
<td></td>
</tr>
</tbody>
</table>

#### Redis 3.2.10 Parameter Changes

**Parameter group family:** redis3.2

ElastiCache for Redis 3.2.10 there are no additional parameters supported.

#### Redis 3.2.6 Parameter Changes

**Parameter group family:** redis3.2

For Redis 3.2.6 there are no additional parameters supported.

#### Redis 3.2.4 Parameter Changes

**Parameter group family:** redis3.2

Beginning with Redis 3.2.4 there are two default parameter groups.

- `default.redis3.2` – When running Redis 3.2.4, specify this parameter group or one derived from it, if you want to create a Redis (cluster mode disabled) replication group and still use the additional features of Redis 3.2.4.
- `default.redis3.2.cluster.on` – Specify this parameter group or one derived from it, when you want to create a Redis (cluster mode enabled) replication group.

**Topics**

- [New Parameters for Redis 3.2.4](#)
- [Parameters Changed in Redis 3.2.4 (Enhanced)](#)

#### New Parameters for Redis 3.2.4

**Parameter group family:** redis3.2

For Redis 3.2.4 the following additional parameters are supported.
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
</table>
| list-max-ziplist-size | Default: -2  
Type: integer  
Modifiable: No | Lists are encoded in a special way to save space. The number of entries allowed per internal list node can be specified as a fixed maximum size or a maximum number of elements. For a fixed maximum size, use -5 through -1, meaning:  
- -5: max size: 64 Kb - not recommended for normal workloads  
- -4: max size: 32 Kb - not recommended  
- -3: max size: 16 Kb - not recommended  
- -2: max size: 8 Kb - recommended  
- -1: max size: 4 Kb - recommended  
- Positive numbers mean store up to exactly that number of elements per list node. |
| list-compress-depth | Default: 0  
Type: integer  
Modifiable: Yes  
Changes Take Effect: Immediately | Lists may also be compressed. Compress depth is the number of quicklist ziplist nodes from each side of the list to exclude from compression. The head and tail of the list are always uncompressed for fast push and pop operations. Settings are:  
- 0: Disable all compression.  
- 1: Start compressing with the 1st node in from the head and tail.  
  `[head]`->`node`->`node`->`...->`node`->`[tail]  
  All nodes except `[head]` and `[tail]` compress.  
- 2: Start compressing with the 2nd node in from the head and tail.  
  `[head]`->`[next]`->`node`->`node`->`...->`node`->`[prev]`->`[tail]  
- Etc. |
| cluster-enabled     | Default: no/yes *  
Type: string  
Modifiable: Yes | Indicates whether this is a Redis (cluster mode enabled) replication group in cluster mode (yes) or a Redis (cluster mode enabled) replication group in non-cluster mode (no). Redis (cluster mode enabled) replication groups in cluster mode can partition their data across up to 90 node groups.  
* Redis 3.2.x has two default parameter groups.  
- `default.redis3.2` – default value no.  
- `default.redis3.2.cluster.on` – default value yes. |
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>cluster-require-full-coverage</code></td>
<td>Default: no Type: boolean Modifiable: yes Changes Take Effect: Immediately</td>
<td>When set to yes, Redis (cluster mode enabled) nodes in cluster mode stop accepting queries if they detect there is at least one hash slot uncovered (no available node is serving it). This way if the cluster is partially down, the cluster becomes unavailable. It automatically becomes available again as soon as all the slots are covered again. However, sometimes you want the subset of the cluster which is working to continue to accept queries for the part of the key space that is still covered. To do so, just set the <code>cluster-require-full-coverage</code> option to no.</td>
</tr>
<tr>
<td><code>hll-sparse-max-bytes</code></td>
<td>Default: 3000 Type: integer Modifiable: Yes Changes Take Effect: Immediately</td>
<td>HyperLogLog sparse representation bytes limit. The limit includes the 16 byte header. When a HyperLogLog using the sparse representation crosses this limit, it is converted into the dense representation. A value greater than 16000 is not recommended, because at that point the dense representation is more memory efficient. We recommend a value of about 3000 to have the benefits of the space-efficient encoding without slowing down PFADD too much, which is O(N) with the sparse encoding. The value can be raised to ~10000 when CPU is not a concern, but space is, and the data set is composed of many HyperLogLogs with cardinality in the 0 - 15000 range.</td>
</tr>
<tr>
<td><code>reserved-memory-percent</code></td>
<td>Default: 25 Type: integer Modifiable: Yes Changes Take Effect: Immediately</td>
<td>The percent of a node's memory reserved for nondata use. By default, the Redis data footprint grows until it consumes all of the node's memory. If this occurs, then node performance will likely suffer due to excessive memory paging. By reserving memory, you can set aside some of the available memory for non-Redis purposes to help reduce the amount of paging. This parameter is specific to ElastiCache, and is not part of the standard Redis distribution. For more information, see reserved-memory and Managing Reserved Memory (p. 466).</td>
</tr>
</tbody>
</table>

**Parameters Changed in Redis 3.2.4 (Enhanced)**

**Parameter group family: redis3.2**

For Redis 3.2.4 the following parameters were changed.
<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>activerehashing</td>
<td>Modifiable: Yes</td>
<td>Modifiable was No.</td>
</tr>
<tr>
<td>databases</td>
<td>Modifiable: Yes if the parameter group is not associated with any cache clusters. Otherwise, no.</td>
<td>Modifiable was No.</td>
</tr>
<tr>
<td>appendonly</td>
<td>Default: off</td>
<td>If you want to upgrade from an earlier Redis version, you must first turn appendonly off.</td>
</tr>
<tr>
<td>appendfsync</td>
<td>Default: off</td>
<td>If you want to upgrade from an earlier Redis version, you must first turn appendfsync off.</td>
</tr>
<tr>
<td>repl-timeout</td>
<td>Default: 60</td>
<td>Is now unmodifiable with a default of 60.</td>
</tr>
<tr>
<td>tcp-keepalive</td>
<td>Default: 300</td>
<td>Default was 0.</td>
</tr>
<tr>
<td>list-max-ziplist-entries</td>
<td>Parameter is no longer available.</td>
<td></td>
</tr>
<tr>
<td>list-max-ziplist-value</td>
<td>Parameter is no longer available.</td>
<td></td>
</tr>
</tbody>
</table>

**Redis 2.8.24 (Enhanced) Added Parameters**

**Parameter group family:** redis2.8

For Redis 2.8.24 there are no additional parameters supported.

**Redis 2.8.23 (Enhanced) Added Parameters**

**Parameter group family:** redis2.8

For Redis 2.8.23 the following additional parameter is supported.

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>close-on-slave-write</td>
<td>Default: yes Type: string (yes/no) Modifiable: Yes Changes Take Effect: Immediately</td>
<td>If enabled, clients who attempt to write to a read-only replica will be disconnected.</td>
</tr>
</tbody>
</table>
How close-on-slave-write works

The close-on-slave-write parameter is introduced by Amazon ElastiCache to give you more control over how your cluster responds when a primary node and a read replica node swap roles due to promoting a read replica to primary.

**Before read-replica promotion**

![Diagram](image1)

If the read-replica cluster is promoted to primary for any reason other than a Multi-AZ enabled replication group failing over, the client will continue trying to write to endpoint A. Because endpoint A is now the endpoint for a read-replica, these writes will fail. This is the behavior for Redis before ElastiCache introducing close-on-replica-write and the behavior if you disable close-on-replica-write.

**Read-replica Promoted – writes to old primary fail**

![Diagram](image2)

With close-on-replica-write enabled, any time a client attempts to write to a read-replica, the client connection to the cluster is closed. Your application logic should detect the disconnection, check the DNS table, and reconnect to the primary endpoint, which now would be endpoint B.
When You Might Disable close-on-replica-write

If disabling close-on-replica-write results in writes to the failing cluster, why disable close-on-replica-write?

As previously mentioned, with close-on-replica-write enabled, any time a client attempts to write to a read-replica the client connection to the cluster is closed. Establishing a new connection to the node takes time. Thus, disconnecting and reconnecting as a result of a write request to the replica also affects the latency of read requests that are served through the same connection. This effect remains in place until a new connection is established. If your application is especially read-heavy or very latency-sensitive, you might keep your clients connected to avoid degrading read performance.

Redis 2.8.22 (Enhanced) Added Parameters

Parameter group family: redis2.8

For Redis 2.8.22 there are no additional parameters supported.

Important

- Beginning with Redis version 2.8.22, repl-backlog-size applies to the primary cluster as well as to replica clusters.
- Beginning with Redis version 2.8.22, the repl-timeout parameter is not supported. If it is changed, ElastiCache will overwrite with the default (60s), as we do with appendonly.

The following parameters are no longer supported.

- appendonly
- appendfsync
- repl-timeout

Redis 2.8.21 Added Parameters

Parameter group family: redis2.8

For Redis 2.8.21, there are no additional parameters supported.

Redis 2.8.19 Added Parameters

Parameter group family: redis2.8
For Redis 2.8.19 there are no additional parameters supported.

**Redis 2.8.6 Added Parameters**

**Parameter group family:** redis2.8

For Redis 2.8.6 the following additional parameters are supported.

<table>
<thead>
<tr>
<th>Name</th>
<th>Default</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>min-slaves-max-lag</td>
<td>Default: 10</td>
<td>Type: integer</td>
<td>The number of seconds within which the primary node must receive a ping request from a read replica. If this amount of time passes and the primary does not receive a ping, then the replica is no longer considered available. If the number of available replicas drops below min-slaves-to-write, then the primary will stop accepting writes at that point. If either this parameter or min-slaves-to-write is 0, then the primary node will always accept writes requests, even if no replicas are available.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td>Changes Take Effect: Immediately</td>
<td></td>
</tr>
<tr>
<td>min-slaves-to-write</td>
<td>Default: 0</td>
<td>Type: integer</td>
<td>The minimum number of read replicas which must be available in order for the primary node to accept writes from clients. If the number of available replicas falls below this number, then the primary node will no longer accept write requests. If either this parameter or min-slaves-max-lag is 0, then the primary node will always accept writes requests, even if no replicas are available.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td>Changes Take Effect: Immediately</td>
<td></td>
</tr>
</tbody>
</table>
| notify-keyspace-events | Default: (an empty string) | Type: string | The types of keyspace events that Redis can notify clients of. Each event type is represented by a single letter:  
  - K — Keyspace events, published with a prefix of __keyspace@<db>__  
  - E — Key-event events, published with a prefix of __keyevent@<db>__  
  - g — Generic, non-specific commands such as DEL, EXPIRE, RENAME, etc.  
  - $ — String commands  
  - l — List commands |
<p>|                       | Modifiable: Yes | Changes Take Effect: Immediately |                                                                                                                                                                                                            |</p>
<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• s — Set commands</td>
<td>You can have any combination of these event types. For example, AKE means that Redis can publish notifications of all event types.</td>
</tr>
<tr>
<td></td>
<td>• h — Hash commands</td>
<td>Do not use any characters other than those listed above; attempts to do so will result in error messages.</td>
</tr>
<tr>
<td></td>
<td>• z — Sorted set commands</td>
<td>By default, this parameter is set to an empty string, meaning that keyspace event notification is disabled.</td>
</tr>
<tr>
<td></td>
<td>• x — Expired events (events</td>
<td></td>
</tr>
<tr>
<td></td>
<td>generated every time a key</td>
<td></td>
</tr>
<tr>
<td></td>
<td>expires)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• e — Evicted events (events</td>
<td></td>
</tr>
<tr>
<td></td>
<td>generated when a key is evicted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>for maxmemory)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• A — An alias for $\text{g$\text{ishzxe}}$</td>
<td></td>
</tr>
<tr>
<td>repl-backlog-size</td>
<td>Default: 1048576</td>
<td>The size, in bytes, of the primary node backlog buffer. The backlog is used for recording updates to data at the primary node. When a read replica connects to the primary, it attempts to perform a partial sync ($\text{psync}$), where it applies data from the backlog to catch up with the primary node. If the $\text{psync}$ fails, then a full sync is required. The minimum value for this parameter is 16384.</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: Immediately</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Note</td>
<td>Beginning with Redis 2.8.22, this parameter applies to the primary cluster as well as the read replicas.</td>
</tr>
</tbody>
</table>
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>repl-backlog-ttl</td>
<td>Default: 3600 Type: integer Modifiable: Yes Changes Take Effect: Immediately</td>
<td>The number of seconds that the primary node will retain the backlog buffer. Starting from the time the last replica node disconnected, the data in the backlog will remain intact until repl-backlog-ttl expires. If the replica has not connected to the primary within this time, then the primary will release the backlog buffer. When the replica eventually reconnects, it will have to perform a full sync with the primary. If this parameter is set to 0, then the backlog buffer will never be released.</td>
</tr>
</tbody>
</table>
| repl-timeout      | Default: 60 Type: integer Modifiable: Yes Changes Take Effect: Immediately | Represents the timeout period, in seconds, for:  
- Bulk data transfer during synchronization, from the read replica's perspective  
- Primary node timeout from the replica's perspective  
- Replica timeout from the primary node's perspective |

### Redis 2.6.13 Parameters

**Parameter group family:** redis2.6

Redis 2.6.13 was the first version of Redis supported by ElastiCache. The following table shows the Redis 2.6.13 parameters that ElastiCache supports.

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>activerehashing</td>
<td>Default: yes Type: string (yes/no) Modifiable: Yes Changes take place: At Creation</td>
<td>Determines whether to enable Redis' active rehashing feature. The main hash table is rehashed ten times per second; each rehash operation consumes 1 millisecond of CPU time. This value is set when you create the parameter group. When assigning a new parameter group to a cluster, this value must be the same in both the old and new parameter groups.</td>
</tr>
<tr>
<td>appendonly</td>
<td>Default: no Type: string Modifiable: Yes Changes Take Effect: Immediately</td>
<td>Enables or disables Redis' append only file feature (AOF). AOF captures any Redis commands that change data in the cache, and is used to recover from certain node failures. The default value is no, meaning AOF is turned off. Set this parameter to yes to enable AOF.</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></td>
<td>For more information, see <a href="#">Mitigating Failures</a>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Append Only Files (AOF) is not supported for cache.t1.micro and cache.t2.* nodes. For nodes of this type, the appendonly parameter value is ignored.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>For Multi-AZ replication groups, AOF is not allowed.</td>
</tr>
<tr>
<td>appendfsync</td>
<td>Default: everysec</td>
<td>When appendonly is set to yes, controls how often the AOF output buffer is written to disk:</td>
</tr>
<tr>
<td></td>
<td>Type: string</td>
<td>• no — the buffer is flushed to disk on an as-needed basis.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td>• everysec — the buffer is flushed once per second. This is the default.</td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: Immediately</td>
<td>• always — the buffer is flushed every time that data in the cluster is modified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Appendfsync is not supported for versions 2.8.22 and later.</td>
</tr>
<tr>
<td>client-output-buffer-limit-normal-hard-limit</td>
<td>Default: 0</td>
<td>If a client's output buffer reaches the specified number of bytes, the client will be disconnected. The default is zero (no hard limit).</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: Immediately</td>
<td></td>
</tr>
<tr>
<td>client-output-buffer-limit-normal-soft-limit</td>
<td>Default: 0</td>
<td>If a client's output buffer reaches the specified number of bytes, the client will be disconnected, but only if this condition persists for client-output-buffer-limit-normal-soft-seconds. The default is zero (no soft limit).</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: Immediately</td>
<td></td>
</tr>
<tr>
<td>client-output-buffer-limit-normal-soft-seconds</td>
<td>Default: 0</td>
<td>If a client's output buffer remains at client-output-buffer-limit-normal-soft-limit bytes for longer than this number of seconds, the client will be disconnected. The default is zero (no time limit).</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: Immediately</td>
<td></td>
</tr>
<tr>
<td>client-output-buffer-limit-pubsub-hard-limit</td>
<td>Default: 33554432</td>
<td>For Redis publish/subscribe clients: If a client's output buffer reaches the specified number of bytes, the client will be disconnected.</td>
</tr>
<tr>
<td>Name</td>
<td>Details</td>
<td>Description</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>-------------</td>
</tr>
</tbody>
</table>
| client-output-buffer-limit-pubsub-soft-limit | Default: 8388608  
Type: integer  
Modifiable: Yes  
Changes Take Effect: Immediately | For Redis publish/subscribe clients: If a client's output buffer reaches the specified number of bytes, the client will be disconnected, but only if this condition persists for `client-output-buffer-limit-pubsub-soft-seconds`. |
| client-output-buffer-limit-pubsub-soft-seconds | Default: 60  
Type: integer  
Modifiable: Yes  
Changes Take Effect: Immediately | For Redis publish/subscribe clients: If a client's output buffer remains at `client-output-buffer-limit-pubsub-soft-limit` bytes for longer than this number of seconds, the client will be disconnected. |
| client-output-buffer-limit-slave-hard-limit | Default: For values see Redis Node-Type Specific Parameters (p. 336)  
Type: integer  
Modifiable: No | For Redis read replicas: If a client's output buffer reaches the specified number of bytes, the client will be disconnected. |
| client-output-buffer-limit-slave-soft-limit | Default: For values see Redis Node-Type Specific Parameters (p. 336)  
Type: integer  
Modifiable: No | For Redis read replicas: If a client's output buffer reaches the specified number of bytes, the client will be disconnected, but only if this condition persists for `client-output-buffer-limit-slave-soft-seconds`. |
| client-output-buffer-limit-slave-soft-seconds | Default: 60  
Type: integer  
Modifiable: No | For Redis read replicas: If a client's output buffer remains at `client-output-buffer-limit-slave-soft-limit` bytes for longer than this number of seconds, the client will be disconnected. |
| databases | Default: 16  
Type: integer  
Modifiable: No  
Changes take place: At Creation | The number of logical partitions the databases is split into. We recommend keeping this value low.  
This value is set when you create the parameter group. When assigning a new parameter group to a cluster, this value must be the same in both the old and new parameter groups. |
| hash-max-ziplist-entries | Default: 512  
Type: integer  
Modifiable: Yes  
Changes Take Effect: Immediately | Determines the amount of memory used for hashes. Hashes with fewer than the specified number of entries are stored using a special encoding that saves space. |
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hash-max-ziplist-value</td>
<td>Default: 64 Type: integer</td>
<td>Determines the amount of memory used for hashes. Hashes with entries that are smaller than the specified number of bytes are stored using a special encoding that saves space.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes Changes Take Effect: Immediately</td>
<td></td>
</tr>
<tr>
<td>list-max-ziplist-entries</td>
<td>Default: 512 Type: integer</td>
<td>Determines the amount of memory used for lists. Lists with fewer than the specified number of entries are stored using a special encoding that saves space.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes Changes Take Effect: Immediately</td>
<td></td>
</tr>
<tr>
<td>list-max-ziplist-value</td>
<td>Default: 64 Type: integer</td>
<td>Determines the amount of memory used for lists. Lists with entries that are smaller than the specified number of bytes are stored using a special encoding that saves space.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes Changes Take Effect: Immediately</td>
<td></td>
</tr>
<tr>
<td>lua-time-limit</td>
<td>Default: 5000 Type: integer</td>
<td>The maximum execution time for a Lua script, in milliseconds, before ElastiCache takes action to stop the script. If lua-time-limit is exceeded, all Redis commands will return an error of the form <code>- BUSY</code> Since this state can cause interference with many essential Redis operations, ElastiCache will first issue a <code>SCRIPT KILL</code> command. If this is unsuccessful, ElastiCache will forcibly restart Redis.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td>maxclients</td>
<td>Default: 65000 Type: integer</td>
<td>The maximum number of clients that can be connected at one time.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: No</td>
<td></td>
</tr>
<tr>
<td>maxmemory-policy</td>
<td>Default: volatile-lru Type: string</td>
<td>The eviction policy for keys when maximum memory usage is reached. Valid values are: <code>volatile-lru</code></td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes Changes Take Effect: Immediately</td>
<td>For more information, see Using Redis as an LRU cache.</td>
</tr>
<tr>
<td>maxmemory-samples</td>
<td>Default: 3 Type: integer</td>
<td>For least-recently-used (LRU) and time-to-live (TTL) calculations, this parameter represents the sample size of keys to check. By default, Redis chooses 3 keys and uses the one that was used least recently.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes Changes Take Effect: Immediately</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Details</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>reserved-memory</td>
<td>Default: 0</td>
<td>The total memory, in bytes, reserved for non-data usage. By default, the Redis node will grow until it consumes the node’s <code>maxmemory</code> (see Redis Node-Type Specific Parameters (p. 336)). If this occurs, then node performance will likely suffer due to excessive memory paging. By reserving memory you can set aside some of the available memory for non-Redis purposes to help reduce the amount of paging. This parameter is specific to ElastiCache, and is not part of the standard Redis distribution. For more information, see <code>reserved-memory-percent</code> and Managing Reserved Memory (p. 466).</td>
</tr>
<tr>
<td>set-max-intset-entries</td>
<td>Default: 512</td>
<td>Determines the amount of memory used for certain kinds of sets (strings that are integers in radix 10 in the range of 64 bit signed integers). Such sets with fewer than the specified number of entries are stored using a special encoding that saves space.</td>
</tr>
<tr>
<td>slave-allow-chaining</td>
<td>Default: no</td>
<td>Determines whether a read replica in Redis can have read replicas of its own.</td>
</tr>
<tr>
<td>slowlog-log-slower-than</td>
<td>Default: 10000</td>
<td>The maximum execution time, in microseconds, for commands to be logged by the Redis Slow Log feature.</td>
</tr>
<tr>
<td>slowlog-max-len</td>
<td>Default: 128</td>
<td>The maximum length of the Redis Slow Log.</td>
</tr>
<tr>
<td>tcp-keepalive</td>
<td>Default: 0</td>
<td>If this is set to a nonzero value (N), node clients are polled every N seconds to ensure that they are still connected. With the default setting of 0, no such polling occurs. <strong>Important</strong> Some aspects of this parameter changed in Redis version 3.2.4. See Parameters Changed in Redis 3.2.4 (Enhanced) (p. 325).</td>
</tr>
</tbody>
</table>
Amazon ElastiCache for Redis
ElastiCache for Redis User Guide
Redis Specific Parameters

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>timeout</td>
<td>Default: 0</td>
<td>The number of seconds a node waits before timing out. Values are:</td>
</tr>
<tr>
<td></td>
<td>Type: integer</td>
<td>• 0 – never disconnect an idle client.</td>
</tr>
<tr>
<td></td>
<td>Modifiable: Yes</td>
<td>• 1-19 – invalid values.</td>
</tr>
<tr>
<td></td>
<td>Changes Take Effect: Immediately</td>
<td>• &gt;=20 – the number of seconds a node waits before disconnecting an idle client.</td>
</tr>
<tr>
<td>zset-max-ziplist-entries</td>
<td>Default: 128</td>
<td>Determines the amount of memory used for sorted sets. Sorted sets with fewer than the specified number of elements are stored using a special encoding that saves space.</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: Immediately</td>
<td></td>
</tr>
<tr>
<td>zset-max-ziplist-value</td>
<td>Default: 64</td>
<td>Determines the amount of memory used for sorted sets. Sorted sets with entries that are smaller than the specified number of bytes are stored using a special encoding that saves space.</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: Immediately</td>
<td></td>
</tr>
</tbody>
</table>

**Note**
If you do not specify a parameter group for your Redis 2.6.13 cluster, then a default parameter group (default.redis2.6) will be used. You cannot change the values of any parameters in the default parameter group; however, you can always create a custom parameter group and assign it to your cluster at any time.

Redis 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 maxmemory, client-output-buffer-limit-slave-hard-limit, and client-output-buffer-limit-slave-soft-limit parameters for each node type. The value of maxmemory is the maximum number of bytes available to you for use, data and other uses, on the node.

**Note**
The maxmemory parameter cannot be modified.

<table>
<thead>
<tr>
<th>Node Type</th>
<th>maxmemory</th>
<th>client-output-buffer-limit-slave-hard-limit</th>
<th>client-output-buffer-limit-slave-soft-limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>cache.t1.micro</td>
<td>142606336</td>
<td>14260633</td>
<td>14260633</td>
</tr>
<tr>
<td>cache.t2.micro</td>
<td>581959680</td>
<td>58195968</td>
<td>58195968</td>
</tr>
<tr>
<td>cache.t2.small</td>
<td>1665138688</td>
<td>166513868</td>
<td>166513868</td>
</tr>
<tr>
<td>cache.t2.medium</td>
<td>3461349376</td>
<td>346134937</td>
<td>346134937</td>
</tr>
<tr>
<td>cache.t3.micro</td>
<td>536870912</td>
<td>53687091</td>
<td>53687091</td>
</tr>
<tr>
<td>cache.t3.small</td>
<td>1471026299</td>
<td>147102629</td>
<td>147102629</td>
</tr>
<tr>
<td>Node Type</td>
<td>maxmemory</td>
<td>client-output-buffer-limit-slave-hard-limit</td>
<td>client-output-buffer-limit-slave-soft-limit</td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>--------------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>cache.t3.medium</td>
<td>3317862236</td>
<td>331786223</td>
<td>331786223</td>
</tr>
<tr>
<td>cache.m1.small</td>
<td>943718400</td>
<td>94371840</td>
<td>94371840</td>
</tr>
<tr>
<td>cache.m1.medium</td>
<td>3093299200</td>
<td>309329920</td>
<td>309329920</td>
</tr>
<tr>
<td>cache.m1.large</td>
<td>7025459200</td>
<td>702545920</td>
<td>702545920</td>
</tr>
<tr>
<td>cache.m1.xlarge</td>
<td>14889779200</td>
<td>1488977920</td>
<td>1488977920</td>
</tr>
<tr>
<td>cache.m2.xlarge</td>
<td>17091788800</td>
<td>1709178880</td>
<td>1709178880</td>
</tr>
<tr>
<td>cache.m2.2xlarge</td>
<td>35022438400</td>
<td>3502243840</td>
<td>3502243840</td>
</tr>
<tr>
<td>cache.m2.4xlarge</td>
<td>70883737600</td>
<td>7088373760</td>
<td>7088373760</td>
</tr>
<tr>
<td>cache.m3.medium</td>
<td>29884416000</td>
<td>3093299200</td>
<td>3093299200</td>
</tr>
<tr>
<td>cache.m3.large</td>
<td>65011712000</td>
<td>6501171200</td>
<td>6501171200</td>
</tr>
<tr>
<td>cache.m3.xlarge</td>
<td>14260633600</td>
<td>1426063360</td>
<td>1426063360</td>
</tr>
<tr>
<td>cache.m3.2xlarge</td>
<td>29989273600</td>
<td>2998927360</td>
<td>2998927360</td>
</tr>
<tr>
<td>cache.m4.large</td>
<td>6892593152</td>
<td>689259315</td>
<td>689259315</td>
</tr>
<tr>
<td>cache.m4.xlarge</td>
<td>15328501760</td>
<td>1532850176</td>
<td>1532850176</td>
</tr>
<tr>
<td>cache.m4.2xlarge</td>
<td>31889126359</td>
<td>3188912636</td>
<td>3188912636</td>
</tr>
<tr>
<td>cache.m4.4xlarge</td>
<td>65257290629</td>
<td>6525729063</td>
<td>6525729063</td>
</tr>
<tr>
<td>cache.m4.10xlarge</td>
<td>166047614239</td>
<td>16604761424</td>
<td>16604761424</td>
</tr>
<tr>
<td>cache.m5.large</td>
<td>6854542746</td>
<td>685454275</td>
<td>685454275</td>
</tr>
<tr>
<td>cache.m5.xlarge</td>
<td>13891921715</td>
<td>1389192172</td>
<td>1389192172</td>
</tr>
<tr>
<td>cache.m5.2xlarge</td>
<td>27966669210</td>
<td>2796666921</td>
<td>2796666921</td>
</tr>
<tr>
<td>cache.m5.4xlarge</td>
<td>56116178125</td>
<td>5611617812</td>
<td>5611617812</td>
</tr>
<tr>
<td>cache.m5.12xlarge</td>
<td>168715971994</td>
<td>16871597199</td>
<td>16871597199</td>
</tr>
<tr>
<td>cache.m5.24xlarge</td>
<td>337500562842</td>
<td>33750056284</td>
<td>33750056284</td>
</tr>
<tr>
<td>cache.c1.xlarge</td>
<td>65011712000</td>
<td>6501171200</td>
<td>6501171200</td>
</tr>
<tr>
<td>cache.r3.large</td>
<td>14470348800</td>
<td>1468006400</td>
<td>1468006400</td>
</tr>
<tr>
<td>cache.r3.xlarge</td>
<td>30513561600</td>
<td>3040870400</td>
<td>3040870400</td>
</tr>
<tr>
<td>cache.r3.2xlarge</td>
<td>62495129600</td>
<td>6081740800</td>
<td>6081740800</td>
</tr>
<tr>
<td>cache.r3.4xlarge</td>
<td>126458265600</td>
<td>12268339200</td>
<td>12268339200</td>
</tr>
<tr>
<td>cache.r3.8xlarge</td>
<td>254384537600</td>
<td>24536678400</td>
<td>24536678400</td>
</tr>
<tr>
<td>cache.r4.large</td>
<td>13201781556</td>
<td>1320178155</td>
<td>1320178155</td>
</tr>
</tbody>
</table>
### Redis Specific Parameters

<table>
<thead>
<tr>
<th>Node Type</th>
<th>maxmemory</th>
<th>client-output-buffer-limit-slave-hard-limit</th>
<th>client-output-buffer-limit-slave-soft-limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>cache.r4.xlarge</td>
<td>26898228839</td>
<td>2689822883</td>
<td>2689822883</td>
</tr>
<tr>
<td>cache.r4.2xlarge</td>
<td>54197537997</td>
<td>5419753799</td>
<td>5419753799</td>
</tr>
<tr>
<td>cache.r4.4xlarge</td>
<td>108858546586</td>
<td>10885854658</td>
<td>10885854658</td>
</tr>
<tr>
<td>cache.r4.8xlarge</td>
<td>218255432090</td>
<td>21825543209</td>
<td>21825543209</td>
</tr>
<tr>
<td>cache.r4.16xlarge</td>
<td>437021573120</td>
<td>43702157312</td>
<td>43702157312</td>
</tr>
<tr>
<td>cache.r5.large</td>
<td>14037181030</td>
<td>1403718103</td>
<td>1403718103</td>
</tr>
<tr>
<td>cache.r5.xlarge</td>
<td>28261849702</td>
<td>2826184970</td>
<td>2826184970</td>
</tr>
<tr>
<td>cache.r5.2xlarge</td>
<td>56711183565</td>
<td>5671118356</td>
<td>5671118356</td>
</tr>
<tr>
<td>cache.r5.4xlarge</td>
<td>113609865216</td>
<td>11360986522</td>
<td>11360986522</td>
</tr>
<tr>
<td>cache.r5.12xlarge</td>
<td>341206346547</td>
<td>34120634655</td>
<td>34120634655</td>
</tr>
<tr>
<td>cache.r5.24xlarge</td>
<td>682485973811</td>
<td>68248597381</td>
<td>68248597381</td>
</tr>
</tbody>
</table>

**Note**

All current generation instance types are created in an Amazon Virtual Private Cloud VPC by default.

T1 instances do not support Multi-AZ with automatic failover.

T1 and T2 instances do not support Redis AOF.

Redis configuration variables `appendonly` and `appendfsync` are not supported on Redis version 2.8.22 and later.
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. 339)
- Internetwork Traffic Privacy (p. 357)
- Identity and Access Management in Amazon ElastiCache (p. 390)
- Logging and Monitoring in ElastiCache (p. 411)
- Compliance Validation for Amazon ElastiCache (p. 449)
- Resilience in Amazon ElastiCache (p. 454)
- Infrastructure Security in AWS Elasticache (p. 457)

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.

Topics
• Data Security in Amazon ElastiCache (p. 340)

Data Security in Amazon ElastiCache

To help keep your data secure, Amazon ElastiCache and Amazon EC2 provide mechanisms to guard against unauthorized access of your data on the server.

Amazon ElastiCache for Redis also provides optional encryption features for data on clusters running Redis versions 3.2.6, 4.0.10 or later:

• In-transit encryption encrypts your data whenever it is moving from one place to another, such as between nodes in your cluster or between your cluster and your application.
• At-rest encryption encrypts your on-disk data during sync and backup operations.

If you want to enable in-transit or at-rest encryption, you must meet the following conditions.

• Your cluster or replication group must be running Redis 3.2.6, 4.0.10 or later.
• Your cluster or replication group must be created in a VPC based on Amazon VPC.
• Optionally, you can also use AUTH and the AUTH token (password) needed to perform operations on this cluster or replication group.
ElastiCache for Redis In-Transit Encryption (TLS)

To help keep your data secure, Amazon ElastiCache and Amazon EC2 provide mechanisms to guard against unauthorized access of your data on the server. By providing in-transit encryption capability, ElastiCache gives you a tool you can use to help protect your data when it is moving from one location to another. For example, you might move data from a primary node to a read replica node within a replication group, or between your replication group and your application.

In-transit encryption is optional and can only be enabled on Redis replication groups when they are created. You enable in-transit encryption on a replication group by setting the parameter `TransitEncryptionEnabled` (CLI: `--transit-encryption-enabled`) to `true` when you create the replication group. You can do this whether you are creating the replication group using the AWS Management Console, the AWS CLI, or the ElastiCache API. If you enable in-transit encryption, you must also provide a value for `CacheSubnetGroup`.

**Important**
The parameters `TransitEncryptionEnabled` (CLI: `--transit-encryption-enabled`) are only available when using the `create-replication-group` operation.

**Topics**
- In-Transit Encryption Overview (p. 342)
- In-Transit Encryption Conditions (p. 342)
- Enabling In-Transit Encryption (p. 342)
- Connecting to Amazon ElastiCache for Redis Nodes Enabled with In-Transit Encryption Using redis-cli (p. 345)
In-Transit Encryption Overview

Amazon ElastiCache in-transit encryption is an optional feature that allows you to increase the security of your data at its most vulnerable points—when it is in transit from one location to another. Because there is some processing needed to encrypt and decrypt the data at the endpoints, enabling in-transit encryption can have some performance impact. You should benchmark your data with and without in-transit encryption to determine the performance impact for your use cases.

ElastiCache in-transit encryption implements the following features:

- **Encrypted connections**—both the server and client connections are Secure Socket Layer (SSL) encrypted.
- **Encrypted replication**—data moving between a primary node and replica nodes is encrypted.
- **Server authentication**—clients can authenticate that they are connecting to the right server.
- **Client authentication**—using the Redis AUTH feature, the server can authenticate the clients.

In-Transit Encryption Conditions

The following constraints on Amazon ElastiCache in-transit encryption should be kept in mind when you plan your implementation:

- In-transit encryption is supported on replication groups running Redis versions 3.2.6, 4.0.10 and later.
- In-transit encryption is supported only for replication groups running in an Amazon VPC.
- In-transit encryption is only supported for replication groups running the following node types.
  - R5, R4, R3
  - M5, M4, M3
  - T2

For more information, see Supported Node Types (p. 65).

- In-transit encryption is enabled by explicitly setting the parameter `TransitEncryptionEnabled` to `true`.
- You can enable in-transit encryption on a replication group only when creating the replication group. You cannot toggle in-transit encryption on and off by modifying a replication group. For information on implementing in-transit encryption on an existing replication group, see Enabling In-Transit Encryption (p. 342).
- To connect to an in-transit encryption enabled replication group, a database must be enabled for transport layer security (TLS). To connect to a replication group that is not in-transit encryption enabled, the database cannot be TLS-enabled.

Because of the processing required to encrypt and decrypt the data at the endpoints, implementing in-transit encryption can reduce performance. Benchmark in-transit encryption compared to no encryption on your own data to determine its impact on performance for your implementation.

**Tip**

Because creating new connections can be expensive, you can reduce the performance impact of in-transit encryption by persisting your SSL connections.

Enabling In-Transit Encryption

You can enable in-transit encryption when you create an ElastiCache for Redis replication group using the AWS Management Console, the AWS CLI, or the ElastiCache API.
Enabling In-Transit Encryption on an Existing Cluster

You can only enable in-transit encryption when you create a Redis replication group. If you have an existing replication group on which you want to enable in-transit encryption, do the following.

To enable in-transit encryption for an existing Redis replication group

1. Create a manual backup of the replication group. For more information, see Making Manual Backups (p. 218).
2. Create a new replication group by restoring from the backup setting the engine version to 3.2.6, 4.0.10 and later, and the parameter `TransitEncryptionEnabled` to `true` (CLI: `--transit-encryption-enabled`). For more information, see Restoring From a Backup with Optional Cluster Resizing (p. 239).
3. Update the endpoints in your application to the new replication group's endpoints. For more information, see Finding Connection Endpoints (p. 204).
4. Delete the old replication group. For more information, see the following:
   - Deleting a Cluster (p. 117)
   - Deleting a Replication Group (p. 190)

Enabling In-Transit Encryption Using the AWS Management Console

To enable in-transit encryption when creating a replication group using the AWS Management Console, make the following selections:

- Choose Redis as your engine.
- Choose engine version 3.2.6, 4.0.10 or later.
- Choose Yes from the Encryption in-transit list.

For the step-by-step process, see the following:

- Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79)
- Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 83)

Enabling In-Transit Encryption Using the AWS CLI

To enable in-transit encryption when creating a Redis replication group using the AWS CLI, use the parameter `transit-encryption-enabled`.

**Enabling In-Transit Encryption on Redis (Cluster Mode Disabled) Cluster (CLI)**

Use the AWS CLI operation `create-replication-group` and the following parameters to create a Redis replication group with replicas that has in-transit encryption enabled:

**Key Parameters:**

- `--engine`—Must be `redis`.
- `--engine-version`—Must be 3.2.6, 4.0.10 or later.
- `--transit-encryption-enabled`—Required. If you enable in-transit encryption you must also provide a value for the `--cache-subnet-group` parameter.
• `--num-cache-clusters`—Must be at least 1. The maximum value for this parameter is six.

For more information, see the following:
• Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch (AWS CLI) (p. 169)
• create-replication-group

Enabling In-Transit Encryption on a Cluster for Redis (Cluster Mode Enabled) (CLI)

Use the AWS CLI operation `create-replication-group` and the following parameters to create a Redis (cluster mode enabled) replication group that has in-transit encryption enabled:

**Key Parameters:**

• `--engine`—Must be `redis`.
• `--engine-version`—Must be 3.2.6, 4.0.10 or later.
• `--transit-encryption-enabled`—Required. If you enable in-transit encryption you must also provide a value for the `--cache-subnet-group` parameter.
• Use one of the following parameter sets to specify the configuration of the replication group's node groups:
  • `--num-node-groups`—Specifies the number of shards (node groups) in this replication group. The maximum value of this parameter is 90.
  • `--replicas-per-node-group`—Specifies the number of replica nodes in each node group. The value specified here is applied to all shards in this replication group. The maximum value of this parameter is 5.
  • `--node-group-configuration`—Specifies the configuration of each shard independently.

For more information, see the following:
• Creating a Redis (Cluster Mode Enabled) Replication Group from Scratch (AWS CLI) (p. 175)
• create-replication-group

Enabling In-Transit Encryption Using the AWS API

To enable in-transit encryption when creating a Redis replication group using the ElastiCache API, set the parameter `TransitEncryptionEnabled` to `true` with either `CreateCacheCluster` for a single node Redis replication group, or `CreateReplicationGroup` for a replication group with read replicas.

Enabling In-Transit Encryption on a Cluster for Redis (Cluster Mode Disabled) (API)

Use the ElastiCache API operation `CreateReplicationGroup` and the following parameters to create a Redis (cluster mode disabled) replication group that has in-transit encryption enabled:

**Key Parameters**

• `Engine`—Must be `redis`.
• `EngineVersion`—Must be 3.2.6, 4.0.10 or later.
• `TransitEncryptionEnabled`—Must set to `true`.

When `TransitEncryptionEnabled` is set to `true`, you must also provide a value for `CacheSubnetGroup`.
• `NumCacheClusters`—Must be at least 1. The maximum value for this parameter is six.
For more information, see the following:

- Creating a Redis (cluster mode disabled) Replication Group from Scratch (ElastiCache API) (p. 172)
- CreateReplicationGroup

**Enabling In-Transit Encryption on a Cluster for Redis (Cluster Mode Enabled) (API)**

Use the ElastiCache API operation `CreateReplicationGroup` and the following parameters to create a Redis (cluster mode enabled) replication group that has in-transit encryption enabled:

**Key Parameters**

- **Engine**—Must be `redis`.
- **EngineVersion**—Must be 3.2.6, 4.0.10 or later.
- **TransitEncryptionEnabled**—Must set to `true`.

When `TransitEncryptionEnabled` is set to `true`, you must also provide a value for `CacheSubnetGroup`.

- Use one of the following parameter sets to specify the configuration of the replication group's node groups:
  - **NumNodeGroups**—Specifies the number of shards (node groups) in this replication group. The maximum value of this parameter is 90 but can be increased to a maximum of 250 via service limit increase request. For more information, see AWS Service Limits.
  - **ReplicasPerNodeGroup**—Specifies the number of replica nodes in each node group. The value specified here is applied to all shards in this replication group. The maximum value of this parameter is 5.
  - **NodeGroupConfiguration**—Specifies the configuration of each shard independently.

For more information, see the following:

- Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (ElastiCache API) (p. 179)
- CreateReplicationGroup

**Connecting to Amazon ElastiCache for Redis Nodes Enabled with In-Transit Encryption Using redis-cli**

To access data from ElastiCache for Redis nodes enabled with in-transit encryption, you use clients that work with Secure Socket Layer (SSL). However, redis-cli doesn't support SSL or Transport Layer Security (TLS).

To work around this, you can use the `stunnel` command to create an SSL tunnel to the redis nodes. You then use redis-cli to connect to the tunnel to access data from encrypted Redis nodes.

**To use redis-cli to connect to a Redis cluster enabled with in-transit encryption**

1. From an SSH client, install `stunnel`.

   ```
   sudo yum install stunnel
   ```

2. Run the following command to create and edit file `'/etc/stunnel/redis-cli.conf'` simultaneously to add a ElastiCache for Redis cluster endpoint to one or more connection parameters, using provided output below as template:
In this example, the config file has two connections, the `redis-cli` and the `redis-cli-replica`. The parameters are set as follows:

- **client** is set to yes to specify this stunnel instance is a client.
- **accept** is set to the client IP. In this example, the master is set to the Redis default 127.0.0.1 on port 6379. The replica must call a different port and set to 6380. You can use ephemeral ports 1024–65535. For more information, see Ephemeral Ports in the Amazon VPC User Guide.
- **connect** is set to the Redis server endpoint. For more information, see Finding Connection Endpoints (p. 204).

3. Start `stunnel`.

```bash
cat /etc/stunnel/redis-cli.conf
```

```ini
fips = no
setuid = root
setgid = root
pid = /var/run/stunnel.pid
debug = 7
delay = yes
options = NO_SSLv2
options = NO_SSLv3
[redis-cli]
  client = yes
  accept = 127.0.0.1:6379
  connect = master.ssltest.wif01h.use1.cache.amazonaws.com:6379
[redis-cli-replica]
  client = yes
  accept = 127.0.0.1:6380
  connect = ssltest-02.ssltest.wif01h.use1.cache.amazonaws.com:6379
```

Use the `netstat` command to confirm that the tunnels started.

```bash
sudo stunnel /etc/stunnel/redis-cli.conf
```

```bash
sudo netstat -tulnp | grep -i stunnel
```

```text
tcp 0 0 127.0.0.1:6379 0.0.0.0:* LISTEN
3189/stunnel
tcp 0 0 127.0.0.1:6380 0.0.0.0:* LISTEN
3189/stunnel
```

4. Connect to the encrypted Redis node using the local endpoint of the tunnel.

- If no AUTH password was used during ElastiCache for Redis cluster creation, this example uses the `redis-cli` to connect to the ElastiCache for Redis server using complete path for `redis-cli`, on Amazon Linux:

  ```bash
  /home/ec2-user/redis-stable/src/redis-cli -h localhost -p 6379
  ```

- If AUTH password was used during Redis cluster creation, this example uses `redis-cli` to connect to the Redis server using complete path for `redis-cli`, on Amazon Linux:

  ```bash
  /home/ec2-user/redis-stable/src/redis-cli -h localhost -p 6379 -a my-secret-password
  ```

  OR
• Change directory to redis-stable and do the following:

If no AUTH password was used during ElastiCache for Redis cluster creation, this example uses the redis-cli to connect to the ElastiCache for Redis server using complete path for redis-cli, on Amazon Linux:

```
src/redis-cli -h localhost -p 6379
```

If AUTH password was used during Redis cluster creation, this example uses redis-cli to connect to the Redis server using complete path for redis-cli, on Amazon Linux:

```
src/redis-cli -h localhost -p 6379 -a my-secret-password
```

This example uses Telnet to connect to the Redis server.

```
telnet localhost 6379
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^]'.
auth MySecretPassword
+OK
get foo
$3
bar
```

5. To stop and close the SSL tunnels, `pkill` the stunnel process.

```
sudo pkill stunnel
```

See Also

- At-Rest Encryption in ElastiCache for Redis (p. 348)
- Authenticating Users with the Redis AUTH Command (p. 354)
- Amazon VPCs and ElastiCache Security (p. 358)
- Identity and Access Management in Amazon ElastiCache (p. 390)
At-Rest Encryption in ElastiCache for Redis

To help keep your data secure, Amazon ElastiCache and Amazon S3 provide different ways to restrict access to data in your cache. For more information, see Amazon VPCs and ElastiCache Security (p. 358) and Identity and Access Management in Amazon ElastiCache (p. 390).

ElastiCache for Redis at-rest encryption is an optional feature to increase data security by encrypting on-disk data. When enabled on a replication group, it encrypts the following aspects:

- Disk during sync, backup and swap operations
- Backups stored in Amazon S3

ElastiCache for Redis offers default (service managed) encryption at rest, as well as ability to use your own symmetric customer managed customer master keys in AWS Key Management Service (KMS).

At-rest encryption can be enabled on a replication group only when it is created. Because there is some processing needed to encrypt and decrypt the data, enabling at-rest encryption can have a performance impact during these operations. You should benchmark your data with and without at-rest encryption to determine the performance impact for your use cases.

For information on encryption in transit, see ElastiCache for Redis In-Transit Encryption (TLS) (p. 341)

Topics

- At-Rest Encryption Conditions (p. 348)
- Using Customer Managed CMKs from AWS KMS (p. 349)
- Enabling At-Rest Encryption (p. 350)
- See Also (p. 354)

At-Rest Encryption Conditions

The following constraints on ElastiCache at-rest encryption should be kept in mind when you plan your implementation of ElastiCache encryption at-rest:

- At-rest encryption is supported on replication groups running Redis version 3.2.6, 4.0.10 or later.
- At-rest encryption is supported only for replication groups running in an Amazon VPC.
- At-rest encryption is only supported for replication groups running the following node types.
  - R5, R4, R3
  - M5, M4, M3
  - T2

For more information, see Supported Node Types (p. 65)

- At-rest encryption is enabled by explicitly setting the parameter `AtRestEncryptionEnabled` to `true`.

- You can enable at-rest encryption on a replication group only when creating the replication group. You cannot toggle at-rest encryption on and off by modifying a replication group. For information on implementing at-rest encryption on an existing replication group, see Enabling At-Rest Encryption (p. 350).

- Encryption of data at rest is not available in the cn-north-1 (Beijing) and cn-northwest-1 (Ningxia), ap-northeast-3 (Asia Pacific (Osaka-Local)) regions. Additionally, the option to use customer managed CMK for encryption at rest is not available in AWS GovCloud (us-gov-east-1 and us-gov-west-1) regions.
Implementing at-rest encryption can reduce performance during backup and node sync operations. Benchmark at-rest encryption compared to no encryption on your own data to determine its impact on performance for your implementation.

**Using Customer Managed CMKs from AWS KMS**

ElastiCache for Redis supports symmetric customer managed customer master keys (CMK) for encryption at rest. Customer-managed CMKs are encryption keys that you create, own and manage in your AWS account. For more information, see Customer Master Keys in the AWS Key Management Service Developer Guide. The keys must be created in AWS KMS before they can be used with ElastiCache.

To learn how to create AWS KMS master keys, see Creating Keys in the AWS Key Management Service Developer Guide.

ElastiCache for Redis allows you to integrate with AWS KMS. For more information, see Using Grants in the AWS Key Management Service Developer Guide. No customer action is needed to enable Amazon ElastiCache integration with AWS KMS.

Note that Amazon ElastiCache currently does not support kms:ViaService. Providing/denying access to Amazon ElastiCache using ViaService will have no effect on key permissions.

You can use AWS CloudTrail to track the requests that Amazon ElastiCache sends to AWS Key Management Service on your behalf. All API calls to AWS Key Management Service related to customer managed CMKs have corresponding CloudTrail logs. You can also see the grants that ElastiCache creates by calling the ListGrants KMS API call.

Once a replication group is encrypted using customer managed CMK, all backups for the replication group are encrypted as follows:

- Automatic daily backups are encrypted using the customer managed CMK associated with the cluster.
- Final backup created when replication group is deleted, is also encrypted using the customer managed CMK associated with the replication group.
- Manually created backups are encrypted by default to use the CMK associated with the replication group. You may override this by choosing another customer managed CMK.
- Copying a backup defaults to using customer managed CMK associated with the source backup. You may override this by choosing another customer managed CMK.

**Note**

- Customer managed CMKs cannot be used when exporting backups to your selected Amazon S3 bucket. However, all backups exported to Amazon S3 are encrypted using Server side encryption. You may choose to copy the backup file to a new S3 object and encrypt using a customer managed CMK, copy the file to another S3 bucket that is set up with default encryption using a CMK or change an encryption option in the file itself.
- You can also use customer managed CMKs to encrypt manually-created backups for replication groups that do not use customer managed CMKs for encryption. With this option, the backup file stored in Amazon S3 is encrypted using a CMK, even though the data is not encrypted on the original replication group.

Restoring from a backup allows you to choose from available encryption options, similar to encryption choices available when creating a new replication group. Also consider:

- If you delete the key or disable the key and revoke grants for the key that you used to encrypt a replication group, the replication group becomes irrecoverable. In other words, it cannot be modified or recovered after a hardware failure. AWS KMS deletes master keys only after a waiting period of at
least seven days. After the key is deleted, you can use a different customer managed CMK to create a backup for archival purposes.

- Automatic key rotation preserves the properties of your AWS KMS master keys, so the rotation has no effect on your ability to access your ElastiCache data. Encrypted Amazon ElastiCache replication groups don't support manual key rotation, which involves creating a new master key and updating any references to the old key. To learn more, see Rotating Customer Master Keys in the AWS Key Management Service Developer Guide.

- Encrypting an ElastiCache replication group using CMK requires one grant per replication group. This grant is used throughout the lifespan of the replication group. Additionally, one grant per backup is used during backup creation. This grant is retired once the backup is created.

- For more information on AWS KMS grants and limits, see Limits in the AWS Key Management Service Developer Guide.

### Enabling At-Rest Encryption

You can enable ElastiCache at-rest encryption when you create a Redis replication group by setting the parameter `AtRestEncryptionEnabled` to `true`. You can't enable at-rest encryption on existing replication groups.

You can enable at-rest encryption when you create an ElastiCache for Redis replication group. You can do so using the AWS Management Console, the AWS CLI, or the ElastiCache API.

When creating a replication group, you can pick one of the following options:

- **Default** – This option uses service managed encryption at rest.
- **Customer managed CMK** – This option allows you to provide the Key ID/ARN from AWS KMS for encryption at rest.

To learn how to create AWS KMS master keys, see Create Keys in the AWS Key Management Service Developer Guide.

### Contents

- Enabling At-Rest Encryption on an Existing Redis Cluster (p. 350)
- Enabling At-Rest Encryption Using the AWS Management Console (p. 351)
- Enabling At-Rest Encryption Using the AWS CLI (p. 351)
  - Enabling At-Rest Encryption on a Redis (Cluster Mode Disabled) Cluster (CLI) (p. 351)
  - Enabling At-Rest Encryption on a Cluster for Redis (Cluster Mode Enabled) (CLI) (p. 352)
- Enabling At-Rest Encryption Using the ElastiCache API (p. 353)
  - Enabling At-Rest Encryption on a Redis (Cluster Mode Disabled) Cluster (API) (p. 353)
  - Enabling At-Rest Encryption on a Cluster for Redis (Cluster Mode Enabled) (API) (p. 353)

### Enabling At-Rest Encryption on an Existing Redis Cluster

You can only enable at-rest encryption when you create a Redis replication group. If you have an existing replication group on which you want to enable at-rest encryption, do the following.

To enable at-rest encryption on an existing replication group

1. Create a manual backup of your existing replication group. For more information, see Making Manual Backups (p. 218).
2. Create a new replication group by restoring from the backup. On the new replication group, enable at-rest encryption. For more information, see Restoring From a Backup with Optional Cluster Resizing (p. 239).
3. Update the endpoints in your application to point to the new replication group.
4. Delete the old replication group. For more information, see Deleting a Cluster (p. 117) or Deleting a Replication Group (p. 190).

Enabling At-Rest Encryption Using the AWS Management Console

To enable at-rest encryption when creating a replication group using the AWS Management Console, make the following selections:

- Choose redis as your engine.
- Choose version 3.2.6, 4.0.10 or later as your engine version.
- Choose Yes from the Encryption at-rest list.

For the step-by-step procedure, see the following:

- Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79)
- Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 83)

Enabling At-Rest Encryption Using the AWS CLI

To enable at-rest encryption when creating a Redis cluster using the AWS CLI, use the --at-rest-encryption-enabled parameter when creating a replication group.

Enabling At-Rest Encryption on a Redis (Cluster Mode Disabled) Cluster (CLI)

The following operation creates the Redis (cluster mode disabled) replication group my-classic-rg with three nodes (--num-cache-clusters), a primary and two read replicas. At-rest encryption is enabled for this replication group (--at-rest-encryption-enabled).

The following parameters and their values are necessary to enable encryption on this replication group:

**Key Parameters**

- --engine—Must be redis.
- --engine-version—Must be 3.2.6, 4.0.10 or later.
- --at-rest-encryption-enabled—Required to enable at-rest encryption.

**Example 1: Redis (Cluster Mode Disabled) Cluster with Replicas**

For Linux, macOS, or Unix:

```
aws elasticache create-replication-group \
  --replication-group-id my-classic-rg \
  --replication-group-description "3 node replication group" \
  --cache-node-type cache.m4.large \
  --engine redis \
  --engine-version 4.0.10 \
  --at-rest-encryption-enabled \
  --num-cache-clusters 3 \
  --cache-parameter-group default.redis4.0
```

For Windows:

```
aws elasticache create-replication-group ^
```
Amazon ElastiCache for Redis
ElastiCache for Redis User Guide
Data Security in Amazon ElastiCache

--replication-group-id my-classic-rg
--replication-group-description "3 node replication group"
--cache-node-type cache.m4.large
--engine redis
--engine-version 4.0.10
--at-rest-encryption-enabled
--num-cache-clusters 3
--cache-parameter-group default.redis4.0

For additional information, see the following:

• Creating a Redis (Cluster Mode Disabled) Replication Group from Scratch (AWS CLI) (p. 169)
• create-replication-group

Enabling At-Rest Encryption on a Cluster for Redis (Cluster Mode Enabled) (CLI)

The following operation creates the Redis (cluster mode enabled) replication group my-clustered-rg with three node groups or shards (--num-node-groups). Each has three nodes, a primary and two read replicas (--replicas-per-node-group). At-rest encryption is enabled for this replication group (--at-rest-encryption-enabled).

The following parameters and their values are necessary to enable encryption on this replication group:

Key Parameters

• --engine—Must be redis.
• --engine-version—Must be 3.2.6, 4.0.10 or later.
• --at-rest-encryption-enabled—Required to enable at-rest encryption.
• --cache-parameter-group—Must be default-redis4.0.cluster.on or one derived from it to make this a cluster mode enabled replication group.

Example 2: A Redis (Cluster Mode Enabled) Cluster

For Linux, macOS, or Unix:

aws elasticache create-replication-group
    --replication-group-id my-clustered-rg
    --replication-group-description "redis clustered cluster"
    --cache-node-type cache.m3.large
    --num-node-groups 3
    --replicas-per-node-group 2
    --engine redis
    --engine-version 4.0.10
    --at-rest-encryption-enabled
    --cache-parameter-group default.redis4.0.cluster.on

For Windows:

aws elasticache create-replication-group
    --replication-group-id my-clustered-rg
    --replication-group-description "redis clustered cluster"
    --cache-node-type cache.m3.large
    --num-node-groups 3
    --replicas-per-node-group 2
    --engine redis

API Version 2015-02-02
352
Enabling At-Rest Encryption Using the ElastiCache API

To enable at-rest encryption when creating a Redis replication group using the ElastiCache API, set the parameter `AtRestEncryptionEnabled` to `true` with `CreateReplicationGroup`.

Enabling At-Rest Encryption on a Redis (Cluster Mode Disabled) Cluster (API)

The following operation creates the Redis (cluster mode disabled) replication group `my-classic-rg` with three nodes (`NumCacheClusters`), a primary and two read replicas. At-rest encryption is enabled for this replication group (`AtRestEncryptionEnabled=true`).

The following parameters and their values are necessary to enable encryption on this replication group:

- **Engine**—Must be `redis`.
- **EngineVersion**—Must be 3.2.6, 4.0.10 or later.
- **AtRestEncryptionEnabled**—Required to be `true` to enable at-rest encryption.

**Example 3: A Redis (Cluster Mode Disabled) Cluster with Replicas**

Line breaks are added for ease of reading.

```
https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateReplicationGroup
  &AtRestEncryptionEnabled=true
  &CacheNodeType=cache.m3.large
  &CacheParameterGroup=default.redis4.0
  &Engine=redis
  &EngineVersion=4.0.10
  &NumCacheClusters=3
  &ReplicationGroupDescription=test%20group
  &ReplicationGroupId=my-classic-rg
  &Version=2015-02-02
  &SignatureVersion=4
  &SignatureMethod=HmacSHA256
  &Timestamp=20150202T192317Z
  &X-Amz-Credential=<credential>
```

For additional information, see the following:

- Creating a Redis (cluster mode disabled) Replication Group from Scratch (ElastiCache API) (p. 172)
- CreateReplicationGroup

**Enabling At-Rest Encryption on a Cluster for Redis (Cluster Mode Enabled) (API)**

The following operation creates the Redis (cluster mode enabled) replication group `my-clustered-rg` with three node groups/shards (`NumNodeGroups`), each with three nodes, a primary and two

---

---
read replicas \((\text{ReplicasPerNodeGroup})\). At-rest encryption is enabled for this replication group \((\text{AtRestEncryptionEnabled}=\text{true})\).

The following parameters and their values are necessary to enable encryption on this replication group:

- **Engine**—Must be redis.
- **AtRestEncryptionEnabled**—Required to be true to enable at-rest encryption.
- **EngineVersion**—Must be 3.2.6, 4.0.10 or later.
- **CacheParameterGroup**—Must be default-redis4.0.cluster.on or one derived from it for this to be a Redis (cluster mode enabled) cluster.

**Example 4: A Redis (Cluster Mode Enabled) Cluster**

Line breaks are added for ease of reading.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=CreateReplicationGroup
&AtRestEncryptionEnabled=true
&CacheNodeType=cache.m3.large
&CacheParameterGroup=default.redis4.0.cluster.on
&Engine=redis
&EngineVersion=4.0.10
&NumNodeGroups=3
&ReplicasPerNodeGroup=2
&ReplicationGroupDescription=test%20group
&ReplicationGroupId=my-clustered-rg
&Version=2015-02-02
&SignatureVersion=4
&SignatureMethod=HmacSHA256
&Timestamp=20150202T192317Z
&X-Amz-Credential=<credential>
```

For additional information, see the following:

- Creating a Replication Group in Redis (Cluster Mode Enabled) from Scratch (ElastiCache API) (p. 179)
- CreateReplicationGroup

**See Also**

- Amazon VPCs and ElastiCache Security (p. 358)
- Identity and Access Management in Amazon ElastiCache (p. 390)

**Authenticating Users with the Redis AUTH Command**

Redis authentication tokens enable Redis to require a token (password) before allowing clients to execute commands, thereby improving data security.

**Topics**

- Overview of AUTH in ElastiCache for Redis (p. 355)
- Applying Authentication to an ElastiCache for Redis Cluster (p. 355)
- Modifying the AUTH Token on an Existing ElastiCache for Redis Cluster (p. 356)
- Related Topics (p. 357)
Overview of AUTH in ElastiCache for Redis

When you use Redis AUTH with your ElastiCache for Redis cluster, there are some refinements.

In particular, be aware of these AUTH token constraints when using AUTH with ElastiCache for Redis:

- Tokens must be 16–128 printable characters.
- Nonalphanumeric characters are restricted to (!, &, #, $, ^, <, >, -).
- AUTH can only be enabled for encryption in-transit enabled ElastiCache for Redis clusters.

To set up a strong token, we recommend that you follow a strict token policy, such as requiring the following:

- Tokens must include at least three of the following character types:
  - Uppercase characters
  - Lowercase characters
  - Digits
  - Nonalphanumeric characters (!, &, #, $, ^, <, >, -)
- Tokens must not contain a dictionary word or a slightly modified dictionary word.
- Tokens must not be the same as or similar to a recently used token.

Applying Authentication to an ElastiCache for Redis Cluster

You can require that users enter a token on a token-protected Redis server. To do this, include the parameter --auth-token (API: AuthToken) with the correct token when you create your replication group or cluster. Also include it in all subsequent commands to the replication group or cluster.

The following AWS CLI operation creates a replication group with encryption in transit (TLS) enabled and the AUTH token This-is-a-sample-token. Replace the subnet group sng-test with a subnet group that exists.

**Key Parameters**

- **--engine** – Must be redis.
- **--engine-version** – Must be 3.2.6, 4.0.10, or later.
- **--transit-encryption-enabled** – Required for authentication and HIPAA eligibility.
- **--auth-token** – Required for HIPAA eligibility. This value must be the correct token for this token-protected Redis server.
- **--cache-subnet-group** – Required for HIPAA eligibility.

For Linux, macOS, or Unix:

```bash
aws elasticache create-replication-group \
  --replication-group-id authtestgroup \
  --replication-group-description authtest \
  --engine redis \
  --engine-version 4.0.10 \
  --cache-node-type cache.m4.large \
  --num-node-groups 1 \
  --replicas-per-node-group 2 \
  --cache-parameter-group default.redis3.2.cluster.on \
  --transit-encryption-enabled \
  --auth-token This-is-a-sample-token 
```
Modifying the AUTH Token on an Existing ElastiCache for Redis Cluster

To make it easier to update your authentication, you can modify the **AUTH** token used on an ElastiCache for Redis cluster. You can make this modification if the engine version is 5.0.5 or higher and if ElastiCache for Redis has encryption in transit enabled.

Modifying the auth token supports two strategies: ROTATE and SET. The ROTATE strategy adds an additional **AUTH** token to the server while retaining the previous token. The SET strategy updates the server to support just a single **AUTH** token. Make these modification calls with the `--apply-immediately` parameter to apply changes immediately.

**Rotating the AUTH Token**

To update a Redis server with a new AUTH token, call the `ModifyReplicationGroup` API with the `--auth-token` parameter as the new auth token and the `--auth-token-update-strategy` with the value ROTATE. Once the modification is complete, the cluster will support the previous AUTH token in addition to the one specified in the `auth-token` parameter.

If this modification is performed on a server that already supports two AUTH tokens, the oldest AUTH token will also be removed during this operation, allowing a server to support up to two most recent AUTH tokens at a given time.

At this point, you can proceed by updating the client to use the latest AUTH token. Once the clients are updated, you can use the SET strategy for AUTH token rotation (explained in the following section) to exclusively start using the new token.

The following AWS CLI operation modifies a replication group to rotate the AUTH token *This-is-the-rotated-token*.

For Linux, macOS, or Unix:

```
aws elasticache modify-replication-group
  --replication-group-id authtestgroup
  --auth-token This-is-the-rotated-token
  --auth-token-update-strategy ROTATE
  --apply-immediately
```

For Windows:

```
aws elasticache modify-replication-group
  --replication-group-id authtestgroup
```

API Version 2015-02-02
Setting the AUTH Token

To update a Redis server with two AUTH tokens to support a single AUTH token, call the ModifyReplicationGroup API operation. Call ModifyReplicationGroup with the --auth-token parameter as the new AUTH token and the --auth-token-update-strategy parameter with the value SET. The auth-token parameter must be the same value as the last AUTH token rotated. After the modification is complete, the Redis server supports only the AUTH token specified in the auth-token parameter.

The following AWS CLI operation modifies a replication group to set the AUTH token to This-is-the-set-token.

For Linux, macOS, or Unix:

```bash
aws elasticache modify-replication-group \
  --replication-group-id authtestgroup \
  --auth-token This-is-the-set-token \
  --auth-token-update-strategy SET \
  --apply-immediately
```

For Windows:

```bash
aws elasticache modify-replication-group ^
  --replication-group-id authtestgroup ^
  --auth-token This-is-the-set-token ^
  --auth-token-update-strategy SET ^
  --apply-immediately
```

Enabling Authentication on an Existing ElastiCache for Redis Cluster

To enable authentication on an existing Redis server, call the ModifyReplicationGroup API operation. Call ModifyReplicationGroup with the --auth-token parameter as the new token and the --auth-token-update-strategy parameter with the value ROTATE.

After the modification is complete, the cluster supports the AUTH token specified in the auth-token parameter in addition to supporting connecting without authentication. Enabling authentication is only supported on Redis servers with encryption in transit (TLS) enabled.

Related Topics

- AUTH token on the redis.io website.

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. 358) explains the type of security group you need for your installation.
- Identity and Access Management in Amazon ElastiCache (p. 390) 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. 359)
- Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 363)
- Creating a Virtual Private Cloud (VPC) (p. 369)
- Creating a Cache Subnet Group (p. 371)
- Creating a Cache Cluster in an Amazon VPC (p. 372)
- Creating a Replication Group in an Amazon VPC (p. 373)
- Connecting to a Cluster or Replication Group Running in an Amazon VPC (p. 374)

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. 359)
- Access Patterns for Accessing an ElastiCache Cluster in an Amazon VPC (p. 363)
- Creating a Virtual Private Cloud (VPC) (p. 369)
- Creating a Cache Subnet Group (p. 371)
- Creating a Cache Cluster in an Amazon VPC (p. 372)
- Creating a Replication Group in an Amazon VPC (p. 373)
- Connecting to a Cluster or Replication Group Running in an Amazon VPC (p. 374)
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.

2. 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.

3. 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.

4. 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.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Amazon VPC is an isolated portion of the AWS Cloud that is assigned its own block of IP addresses.</td>
</tr>
<tr>
<td>2</td>
<td>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.</td>
</tr>
<tr>
<td>3</td>
<td>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.</td>
</tr>
<tr>
<td>4</td>
<td>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.</td>
</tr>
</tbody>
</table>
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](https://aws.amazon.com/vpc/).
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. 363)
- Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs (p. 364)
  - Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in the Same Region (p. 365)
  - Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in Different Regions (p. 366)
- Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center (p. 366)
  - Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using VPN Connectivity (p. 367)
  - Accessing an ElastiCache Cluster from an Application Running in a Customer's Data Center Using Direct Connect (p. 368)

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

![Diagram](image)

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 Redis clusters and replication groups is 6379.

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 Redis clusters and replication groups is 6379.
   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. 365)
- Accessing an ElastiCache Cluster when it and the Amazon EC2 Instance are in Different Amazon VPCs in Different Regions (p. 366)
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</td>
<td>igw-bf6cccd8</td>
</tr>
<tr>
<td>172.16.0.0/16</td>
<td>pcx-a894f1c1</td>
<td></td>
<td></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.
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 and State]

   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.
d. 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. 375).

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. 372).
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 Redis cache cluster, see Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79). In step 6.d 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 or Replication Group Running in an Amazon VPC (p. 374).
Creating a Replication Group in an Amazon VPC

In this example, you create a Redis replication group in your Amazon VPC.

Creating a Replication Group in an Amazon VPC (Console)

To launch a Redis (cluster mode disabled) replication group in a VPC, see Creating a Redis (cluster mode disabled) Replication Group from Scratch (p. 169) In step 5.b, select a VPC subnet group.

To launch a Redis (cluster mode enabled) replication group, see Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 175) In step 6.d, select a VPC subnet group.

You have now launched a Redis replication group inside an Amazon VPC. For an example of one way to connect to your new replication group running in the Amazon VPC, continue to Connecting to a Cluster or Replication Group Running in an Amazon VPC (p. 374).
Connecting to a Cluster or Replication Group 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. 29) 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. 28).

**Topics**
- Creating a Subnet Group (p. 375)
- Assigning a Subnet Group to a Cluster or Replication Group (p. 378)
- Modifying a Subnet Group (p. 379)
- Deleting a Subnet Group (p. 381)
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`.

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

```plaintext
https://elasticache.us-west-2.amazonaws.com/
  ?Action=CreateCacheSubnetGroup
  &CacheSubnetGroupDescription=Testing
  &CacheSubnetGroupName=mysubnetgroup
  &SignatureMethod=HmacSHA256
  &SignatureVersion=4
```

API Version 2015-02-02
&SubnetIds.member.1=subnet-53df9c3a
&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
Assigning a Subnet Group to a Cluster or Replication Group

After you have created a subnet group, you can launch a cluster or replication group in an Amazon VPC. For more information, see the following.

- **Standalone Redis cluster** – To launch a single-node Redis cluster, see Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79). In step 5.a (Advanced Redis Settings), choose a VPC subnet group.

- **Redis (cluster mode disabled) replication group** – To launch a Redis (cluster mode disabled) replication group in a VPC, see Creating a Redis (cluster mode disabled) Replication Group from Scratch (p. 169). In step 5.b (Advanced Redis Settings), choose a VPC subnet group.

- **Redis (cluster mode enabled) replication group** – Creating a Redis (Cluster Mode Enabled) Cluster (Console) (p. 175). In step 5.a (Advanced Redis 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:

```
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:

```
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:

```
{
  "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"
  }
}
```
Modifying Subnet Groups (ElastiCache API)

Using the ElastiCache API, call `ModifyCacheSubnetGroup` with the following parameters:

- `CacheSubnetGroupName=my subnet group`
- 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:

```bash
aws elasticache delete-cache-subnet-group --cache-subnet-group-name mysubnetgroup
```

For Windows:

```bash
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.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=DeleteCacheSubnetGroup
&CacheSubnetGroupName=mysubnetgroup
&SignatureMethod=HmacSHA256
&SignatureVersion=4
&Timestamp=20141201T220302Z
&Version=2014-12-01
&X-Amz-Algorithm=AWS4-HMAC-SHA256
```

API Version 2015-02-02
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. 358).

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. 383)
- Listing Available Security Groups (p. 385)
- Viewing a Security Group (p. 387)
- Authorizing Network Access to an Amazon EC2 Security Group (p. 389)
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. 358).

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. 358).

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.

```
{
  "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.

```bash
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeCacheSecurityGroups
&MaxRecords=100
&Version=2015-02-02
```
Amazon ElastiCache for Redis
ElastiCache for Redis User Guide
Security Groups: EC2-Classic

| &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> |
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. 358).

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.

```json
{
  "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. 358).

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:

```
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:

```
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:

```
{
  "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. 391)
- Access Control (p. 392)

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. 393)
- **Using Identity-Based Policies (IAM Policies) for Amazon ElastiCache** (p. 397)
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. 393)
• Understanding Resource Ownership (p. 393)
• Managing Access to Resources (p. 394)
• Specifying Policy Elements: Actions, Effects, Resources, and Principals (p. 395)
• Specifying Conditions in a Policy (p. 395)

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. 394)
- Resource-Based Policies (p. 395)

**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 (*).

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Sid": "DescribeCacheClusters",
    "Effect": "Allow",
    "Action": ["elasticache:DescribeCacheClusters"],
```

API Version 2015-02-02
394
For more information about using identity-based policies with ElastiCache, see Using Identity-Based Policies (IAM Policies) for Amazon ElastiCache (p. 397). 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. 393)), 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. 393).

- **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. 408).

**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. 393).

The sections in this topic cover the following:

- Permissions Required to Use the Amazon ElastiCache Console (p. 398)
- AWS-Managed (Predefined) Policies for Amazon ElastiCache (p. 398)
- Customer-Managed Policy Examples (p. 399)

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:CreateReplicationGroup`, `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. 408).
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. 408).

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. 401)). 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. 398).

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. 399)
- Example 2: Allow a User Read-Only Access to ElastiCache Resources (p. 400)
- Example 3: Allow a User to Perform Common ElastiCache System Administrator Tasks (p. 400)
- Example 4: Allow a User to Access All ElastiCache API Actions (p. 400)
- Example 5: Allow a User to Call IAM CreateServiceLinkedRole API (p. 401)

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.

```json
{
  "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.

```json
{
    "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.

```json
{
    "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.

```json
{
    "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.

```
{
  "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. 402)
- Creating a Service-Linked Role (IAM) (p. 403)
  - Creating a Service-Linked Role (IAM Console) (p. 403)
  - Creating a Service-Linked Role (IAM CLI) (p. 403)
  - Creating a Service-Linked Role (IAM API) (p. 403)
- Editing the Description of a Service-Linked Role for Amazon ElastiCache (p. 404)
  - Editing a Service-Linked Role Description (IAM Console) (p. 404)
  - Editing a Service-Linked Role Description (IAM CLI) (p. 404)
Service-Linked Role Permissions for Amazon ElastiCache

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 AWSServiceRoleForElastiCache 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 AWSServiceRoleForElastiCache 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/AWSServiceRoleForElastiCache*",
  "Condition": {"StringLike": {"iam:AWSServiceName": "elasticache.amazonaws.com"}}
}
```

To allow an IAM entity to delete AWSServiceRoleForElastiCache 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/AWSServiceRoleForElastiCache*",
  "Condition": {"StringLike": {"iam:AWSServiceName": "elasticache.amazonaws.com"}}
}
```
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**

   ```
   $ 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:

   ```
   $ aws iam update-role-description \
   --role-name AWSServiceRoleForElastiCache \ 
   --description "new description"
   ```

   For Windows:

   ```
   $ 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.
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 or replication groups) 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. 117)
  - Using the AWS CLI (p. 117)
  - Using the ElastiCache API (p. 118)
- To delete a replication group, see the following:
  - Deleting a Replication Group (Console) (p. 190)
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.

   ```shell
   $ 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 deletion-task-id from the response to check the status of the deletion task. Enter the following to submit a service-linked role deletion request.

   ```shell
   $ aws iam delete-service-linked-role --role-name role-name
   ```

3. Enter the following to check the status of the deletion task.

   ```shell
   $ 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 role, 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. 392) 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:** *

### CopySnapshot

**Action:** `elasticache:CopySnapshot`

**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: *
CreateReplicationGroup

Action: elasticache:CreateReplicationGroup

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.

CreateSnapshot

Action: elasticache:CreateSnapshot

Resource: *
DeleteCacheCluster

Action: elasticache:DeleteCacheCluster

Resource: *
DeleteCacheParameterGroup

Action: elasticache:DeleteCacheParameterGroup

Resource: *
DeleteCacheSecurityGroup

Action: elasticache:DeleteCacheSecurityGroup

Resource: *
DeleteCacheSubnetGroup

Action: elasticache:DeleteCacheSubnetGroup

Resource: *
DeleteReplicationGroup

Action: elasticache:DeleteReplicationGroup

Resource: *
DeleteSnapshot

Action: elasticache:DeleteSnapshot

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: *
DescribeReplicationGroups

Action: elasticache:DescribeReplicationGroups

Resource: *
DescribeReservedCacheNodes

Action: elasticache:DescribeReservedCacheNodes

Resource: *
DescribeReservedCacheNodesOfferings

Action: elasticache:DescribeReservedCacheNodesOfferings

Resource: *
DescribeSnapshots

Action: elasticache:DescribeSnapshots

Resource: *
ListTagsForResource

Action: elasticache:ListTagsForResource
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. 413)
- Monitoring ElastiCache Events (p. 423)
- Monitoring Costs with Cost Allocation Tags (p. 432)
- Managing Costs with Reserved Nodes (p. 439)
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. 418) in this guide.

Topics

- Host-Level Metrics (p. 413)
- Metrics for Redis (p. 414)
- Which Metrics Should I Monitor? (p. 418)
- Choosing Metric Statistics and Periods (p. 420)
- Monitoring CloudWatch Cluster and Node Metrics (p. 420)

Host-Level Metrics

The AWS/ElastiCache namespace includes the following host-level metrics for individual cache nodes.

See Also

- Metrics for Redis (p. 414)

<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 for the entire host. Because Redis is single-threaded, we recommend you monitor EngineCPUUtilization metric for nodes with 4 or more vCPUs.</td>
<td>Percent</td>
</tr>
<tr>
<td>FreeableMemory</td>
<td>The amount of free memory available on the host. This is derived from the RAM, buffers and cache that the OS reports as freeable.</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 sent out on all network interfaces by the instance.</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</td>
<td>Count</td>
</tr>
</tbody>
</table>
### Metrics for Redis

The AWS/ElastiCache namespace includes the following Redis metrics.

With the exception of ReplicationLag and EngineCPUUtilization, these metrics are derived from the Redis `info` command. Each metric is calculated at the cache node level.

For complete documentation of the Redis `info` command, see [http://redis.io/commands/info](http://redis.io/commands/info).

**See Also**
- [Host-Level Metrics (p. 413)](#)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>ActiveDefragHits</td>
<td>The number of value reallocations per minute performed by the active defragmentation process. This is derived from <code>active_defrag_hits</code> statistic at Redis <code>INFO</code>.</td>
<td>Number</td>
</tr>
<tr>
<td>BytesUsedForCache</td>
<td>The total number of bytes allocated by Redis for all purposes, including the dataset, buffers, etc. This is derived from <code>used_memory</code> statistic at Redis <code>INFO</code>.</td>
<td>Bytes</td>
</tr>
<tr>
<td>CacheHits</td>
<td>The number of successful read-only key lookups in the main dictionary. This is derived from <code>keyspace_hits</code> statistic at Redis <code>INFO</code>.</td>
<td>Count</td>
</tr>
<tr>
<td>CacheMisses</td>
<td>The number of unsuccessful read-only key lookups in the main dictionary. This is derived from <code>keyspace_misses</code> at Redis <code>INFO</code>.</td>
<td>Count</td>
</tr>
<tr>
<td>CurrConnections</td>
<td>The number of client connections, excluding connections from read replicas. uses two to four of the connections to monitor the cluster in each case. This is derived from the <code>connected_clients</code> statistic at Redis <code>INFO</code>.</td>
<td>Count</td>
</tr>
<tr>
<td>EngineCPUUtilization</td>
<td>Provides CPU utilization of the Redis engine thread. Since Redis is single-threaded, you can use this metric to analyze the load of the Redis process itself. The EngineCPUUtilization metric provides a more precise visibility of the Redis process and can be used in conjunction with CPUUtilization metric, which exposes CPU utilization for the server instance as a whole, including other operating system and management processes. For larger node types with 4vCPUs or more, use the</td>
<td>Percent</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
<td>Unit</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>EngineCPUUtilization</td>
<td>EngineCPUUtilization metric to monitor and set thresholds for scaling. For smaller node types with 2vCPUs or less, use the CPUUtilization metric.</td>
<td></td>
</tr>
<tr>
<td>Evictions</td>
<td>The number of keys that have been evicted due to the maxmemory limit. This is derived from the evicted_keys statistic at Redis INFO.</td>
<td>Count</td>
</tr>
<tr>
<td>MasterLinkHealthStatus</td>
<td>This status has two values: 0 or 1. The value 0 indicates that data in the Elasticache primary node is not in sync with Redis on EC2. The value of 1 indicates that the data is in sync. To complete the migration, use the CompleteMigration API.</td>
<td>Boolean</td>
</tr>
<tr>
<td>NewConnections</td>
<td>The total number of connections that have been accepted by the server during this period. This is derived from the total_connections_received statistic at Redis INFO.</td>
<td>Count</td>
</tr>
<tr>
<td>Reclaimed</td>
<td>The total number of key expiration events. This is derived from the expired_keys statistic at Redis INFO.</td>
<td>Count</td>
</tr>
<tr>
<td>ReplicationBytes</td>
<td>For nodes in a replicated configuration, ReplicationBytes reports the number of bytes that the primary is sending to all of its replicas. This metric is representative of the write load on the replication group. This is derived from the master_repl_offset statistic at Redis INFO.</td>
<td>Bytes</td>
</tr>
<tr>
<td>ReplicationLag</td>
<td>This metric is only applicable for a node running as a read replica. It represents how far behind, in seconds, the replica is in applying changes from the primary node.</td>
<td></td>
</tr>
<tr>
<td>SaveInProgress</td>
<td>This binary metric returns 1 whenever a background save (forked or forkless) is in progress, and 0 otherwise. A background save process is typically used during snapshots and syncs. These operations can cause degraded performance. Using the SaveInProgress metric, you can diagnose whether or not degraded performance was caused by a background save process. This is derived from the rdb_bgsave_in_progress statistic at Redis INFO.</td>
<td>Count</td>
</tr>
</tbody>
</table>

**EngineCPUUtilization availability**
Nodes in a region created or replaced after the date and time specified in the following table will include the EngineCPUUtilization metric.

<table>
<thead>
<tr>
<th>Region</th>
<th>Region name</th>
<th>EngineCPUUtilization availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>us-east-2</td>
<td></td>
<td>February 16, 2017 17:21 (UTC)</td>
</tr>
<tr>
<td>us-east-1</td>
<td></td>
<td>February 8, 2017 21:20 (UTC)</td>
</tr>
<tr>
<td>us-west-1</td>
<td></td>
<td>February 14, 2017 22:23 (UTC)</td>
</tr>
<tr>
<td>us-west-2</td>
<td></td>
<td>February 20, 2017 19:20 (UTC)</td>
</tr>
<tr>
<td>ap-northeast-1</td>
<td></td>
<td>February 14, 2017 19:58 (UTC)</td>
</tr>
<tr>
<td>ap-northeast-2</td>
<td></td>
<td>Available on all nodes.</td>
</tr>
<tr>
<td>ap-northeast-3</td>
<td>Asia Pacific (Osaka-Local)</td>
<td>Available on all nodes.</td>
</tr>
<tr>
<td>ap-south-1</td>
<td></td>
<td>February 7, 2017 02:51 (UTC)</td>
</tr>
<tr>
<td>ap-southeast-1</td>
<td></td>
<td>February 13, 2017 23:40 (UTC)</td>
</tr>
<tr>
<td>ap-southeast-2</td>
<td></td>
<td>February 14, 2017 03:33 (UTC)</td>
</tr>
<tr>
<td>ca-central-1</td>
<td></td>
<td>Available on all nodes.</td>
</tr>
<tr>
<td>cn-north-1</td>
<td></td>
<td>February 16, 2017 22:39 (UTC)</td>
</tr>
<tr>
<td>cn-northwest-2</td>
<td></td>
<td>Available on all nodes.</td>
</tr>
<tr>
<td>eu-central-1</td>
<td></td>
<td>February 15, 2017 00:46 (UTC)</td>
</tr>
<tr>
<td>eu-west-1</td>
<td></td>
<td>February 7, 2017 21:30 (UTC)</td>
</tr>
<tr>
<td>eu-west-2</td>
<td></td>
<td>February 16, 2017 18:58 (UTC)</td>
</tr>
<tr>
<td>eu-west-3</td>
<td>EU (Paris)</td>
<td>Available on all nodes.</td>
</tr>
<tr>
<td>sa-east-1</td>
<td></td>
<td>February 7, 2017 04:35 (UTC)</td>
</tr>
<tr>
<td>us-gov-west-1</td>
<td></td>
<td>February 16, 2017 20:11 (UTC)</td>
</tr>
</tbody>
</table>

These are aggregations of certain kinds of commands, derived from info commandstats. For a full list of available commands, see redis commands.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>CurrItems</td>
<td>The number of items in the cache. This is derived from the Redis keyspace statistic, summing all of the keys in the entire keyspace.</td>
<td>Count</td>
</tr>
<tr>
<td>GetTypeCmds</td>
<td>The total number of read-only type commands. This is derived from the Redis commandstats statistic by summing all of the read-only type commands (get, hget, scard, lrange, etc.)</td>
<td>Count</td>
</tr>
<tr>
<td>HashBasedCmds</td>
<td>The total number of commands that are hash-based. This is derived from the Redis commandstats statistic by summing all of the</td>
<td>Count</td>
</tr>
<tr>
<td>Metric</td>
<td>Description</td>
<td>Unit</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>commands that act upon one or more hashes (hget, hkeys, hvals, hdel, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HyperLogLogBasedCmds</td>
<td>The total number of HyperLogLog-based commands. This is derived from the Redis commandstats statistic by summing all of the pf type of commands (pfadd, pfcount, pfmerge, etc.).</td>
<td>Count</td>
</tr>
<tr>
<td>KeyBasedCmds</td>
<td>The total number of commands that are key-based. This is derived from the Redis commandstats statistic by summing all of the commands that act upon one or more keys across multiple data structures (del, expire, rename, etc.).</td>
<td>Count</td>
</tr>
<tr>
<td>ListBasedCmds</td>
<td>The total number of commands that are list-based. This is derived from the Redis commandstats statistic by summing all of the commands that act upon one or more lists (lindex, lrange, lpush, ltrim, etc.).</td>
<td>Count</td>
</tr>
<tr>
<td>SetBasedCmds</td>
<td>The total number of commands that are set-based. This is derived from the Redis commandstats statistic by summing all of the commands that act upon one or more sets (scard, sdiff, sadd, sunion, etc.).</td>
<td>Count</td>
</tr>
<tr>
<td>SetTypeCmds</td>
<td>The total number of write types of commands. This is derived from the Redis commandstats statistic by summing all of the mutative types of commands that operate on data (set, hset, sadd, lpop, etc.)</td>
<td>Count</td>
</tr>
<tr>
<td>SortedSetBasedCmds</td>
<td>The total number of commands that are sorted set-based. This is derived from the Redis commandstats statistic by summing all of the commands that act upon one or more sorted sets (zcount, zrange, zrank, sadd, etc.).</td>
<td>Count</td>
</tr>
<tr>
<td>StringBasedCmds</td>
<td>The total number of commands that are string-based. This is derived from the Redis commandstats statistic by summing all of the commands that act upon one or more strings (strlen, setex, setrange, etc.).</td>
<td>Count</td>
</tr>
<tr>
<td>StreamBasedCmds</td>
<td>The total number of commands that are stream-based. This is derived from the Redis commandstats statistic by summing all of the commands that act upon one or more streams data types (xrange, xlen, xadd, xdel, etc.).</td>
<td>Count</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. 418)
- EngineCPUUtilization (p. 418)
- SwapUsage (p. 418)
- Evictions (p. 419)
- CurrConnections (p. 419)

CPUUtilization

This is a host-level metric reported as a percentage. For more information, see Host-Level Metrics (p. 413).

For smaller node types with 2vCPUs or less, use the CPUUtilization metric to monitor your workload.

Generally speaking, we suggest you set your threshold at 90% of your available CPU. Because Redis is single-threaded, the actual threshold value should be calculated as a fraction of the node's total capacity. For example, suppose you are using a node type that has two cores. In this case, the threshold for CPUUtilization would be 90/2, or 45%. To find the number of cores (vCPUs) your node type has, see Pricing.

You will need to determine your own threshold, based on the number of cores in the cache node that you are using. If you exceed this threshold, and your main workload is from read requests, scale your cache cluster out by adding read replicas. If the main workload is from write requests, depending on your cluster configuration, we recommend that you:

- **Redis (cluster mode disabled) clusters**: scale up by using a larger cache instance type.
- **Redis (cluster mode enabled) clusters**: add more shards to distribute the write workload across more primary nodes.

Tip

Instead of using the Host-Level metric CPUUtilization, Redis users might be able to use the Redis metric EngineCPUUtilization, which reports the percentage of usage on the Redis engine core. To see if this metric is available on your nodes and for more information, see Metrics for Redis.

EngineCPUUtilization

For larger node types with 4vCPUs or more, you may want to use the EngineCPUUtilization metric, which reports the percentage of usage on the Redis engine core. To see if this metric is available on your nodes and for more information, see Metrics for Redis.

SwapUsage

This is a host-level metric reported in bytes. For more information, see Host-Level Metrics (p. 413).

This metric should not exceed 50 MB. If it does, see the following topics:
- **Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464)**
- **Managing Reserved Memory (p. 466)**
Evictions

This is a cache engine metric. We recommend that you determine your own alarm threshold for this metric based on your application needs.

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. 420).

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

API Version 2015-02-02
420
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=mymachecluster,CacheNodeId=0002

Example

```
http://monitoring.amazonaws.com/
?Action=GetMetricStatistics
&SignatureVersion=4
&Version=2014-12-01
&StartTime=2018-07-05T00:00:00
&EndTime=2018-07-06T00:00:00
&Period=3600
&Statistics.member.1=A
&Dimensions.member.1="CacheClusterId=machecluster"
&Dimensions.member.2="CacheNodeId=0002"
&Namespace=AWS/ElastiCache
&MeasureName=CPUUtilization
&Timestamp=2018-07-07T17%3A48%3A21.746Z
&AWSAccessKeyId=<AWS Access Key ID>
&Signature=<Signature>
```
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. 423)
- Viewing ElastiCache Events (p. 426)
- Event Notifications and Amazon SNS (p. 428)

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. To add an Amazon SNS topic for a replication group, in step 2, instead of choosing a cluster, choose a replication group then follow the same remaining steps.

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:
aws elasticache modify-cache-cluster \   
   --cache-cluster-id my-cluster \   

For Windows:

aws elasticache modify-cache-cluster ^   
   --cache-cluster-id my-cluster ^   

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   
   &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 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 Redis, in the navigation pane choose Redis.  
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 \\n  --cache-cluster-id my-cluster \\n  --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**

```plaintext
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.

```
aws elasticache describe-events --source-type cache-cluster --max-items 40
```

The following code lists all events for the past 24 hours (1440 minutes).

```
aws elasticache describe-events --duration 1440
```

The output from the `describe-events` command looks something like this.

```
{
  "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.

```url
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).

```url
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>
     (...output omitted...)
   </Events>
 </DescribeEventsResult>
</DescribeEventsResponse>
```

For more information, such as available parameters and permitted parameter values, see `DescribeEvents`.

API Version 2015-02-02
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**

```json
{
    "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</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</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>
<tr>
<td>Event Name</td>
<td>Message</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ElastiCache:CacheNodeReplaceComplete</td>
<td>cluster-name</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</td>
<td>cluster-name</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</td>
<td>cluster-name</td>
<td>The deletion of a cache cluster and all associated cache nodes has completed.</td>
</tr>
<tr>
<td>ElastiCache:FailoverComplete (Redis only)</td>
<td>cluster-name</td>
<td>Failover over to a replica node was successful.</td>
</tr>
<tr>
<td>ElastiCache:NodeReplacementCanceled</td>
<td>cluster-name</td>
<td>A node in your cluster that was scheduled for replacement is no longer scheduled for replacement.</td>
</tr>
<tr>
<td>ElastiCache:NodeReplacementRescheduled</td>
<td>cluster-name</td>
<td>A node in your cluster previously scheduled for replacement has been rescheduled for replacement during the new window described in the notification. For information on what actions you can take, see Replacing Nodes (p. 69).</td>
</tr>
</tbody>
</table>
### Event Name | Message | Description
--- | --- | ---
ElastiCache:NodeReplacementScheduled | cluster-name | 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. 69).

ElastiCache:RemoveCacheNodeComplete | cluster-name | A cache node has been removed from the cache cluster.

ElastiCache:SnapshotComplete (Redis only) | cluster-name | A cache snapshot has completed successfully.

ElastiCache:SnapshotFailed (Redis only) | cluster-name | 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.

### Related topics
- Viewing ElastiCache Events (p. 426)
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 Redis nodes. When you add, list, modify, copy, or remove a tag, the operation is applied only to the specified node.

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>`


- 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. 433)
- Managing Your Cost Allocation Tags Using the AWS CLI (p. 434)
- Managing Your Cost Allocation Tags Using the ElastiCache API (p. 437)

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.

Redis clusters may have zero, one, or multiple shards. Because tags are added to Redis nodes rather than the entire cluster, procedure for managing tags on Redis clusters is slightly different for each of the three configurations when using the AWS Management Console. Use one of the two following procedures to manage tags on your Redis cluster.

Managing tags on a Redis cluster with zero shards using the AWS Management Console

2. Choose Redis.
3. Choose the name of the cluster with zero shards you want to add tags to.
4. Choose the box to the left of the cluster's node name.
5. From the Actions list, choose Manage tags, and then use the dialog box to manage your tags.

6. 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.

7. When you're finished, choose **Apply Changes**.

### Managing tags on a Redis cluster with one or more shards using the AWS Management Console


2. Choose **Redis**.

3. Choose the name of the sharded cluster's you want to add tags to.

4. If this cluster has only one shard, skip to step 6. If this cluster has multiple shards, continue at step 5.

5. From the list of shards, choose the name of the shard which has the node you want to add tags to.

6. Choose the box to the left of the node you want to add tags to.

7. From the **Actions** list, choose **Manage tags**, and then use the dialog box to manage your tags.

![Manage Tags](image)

8. 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.

9. 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 Redis nodes. The node to be tagged is specified using an ARN (Amazon Resource Name).


Topics
- Listing Tags Using the AWS CLI (p. 435)
- Adding Tags Using the AWS CLI (p. 435)
- Modifying Tags Using the AWS CLI (p. 436)
- Removing Tags Using the AWS CLI (p. 436)

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 Redis node my-cluster-001 in the my-cluster cluster in region us-west-2.

For Linux, macOS, or Unix:

```
aws elasticache list-tags-for-resource \
```

For Windows:

```
aws elasticache list-tags-for-resource ^
```

Output from this operation will look something like the following, a list of all the tags on the resource.

```
{
   "TagList": [
      {
         "Value": "10110",
         "Key": "CostCenter"
      },
      {
         "Value": "EC2",
         "Key": "Service"
      }
   ]
}
```

If there are no tags on the resource, the output will be an empty TagList.

```
{
   "TagList": []
}
```

For more information, see the AWS CLI for ElastiCache list-tags-for-resource.

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 node my-cluster-001 in 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.

```
{
  "TagList": [
    {
      "Value": "elasticache",
      "Key": "Service"
    },
    {
      "Value": "us-west-2",
      "Key": "Region"
    }
  ]
}
```

For more information, see the AWS CLI for ElastiCache add-tags-to-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 a node in an ElastiCache for Redis 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 node in an ElastiCache for Redis 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 node `my-cluster-001` in the cluster `my-cluster` in the us-west-2 region.

For Linux, macOS, or Unix:

```bash
aws elasticache remove-tags-from-resource \
--tag-keys Service
```

For Windows:

```bash
aws elasticache remove-tags-from-resource ^
--tag-keys 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).

**Sample arn:** `arn:aws:elasticache:us-west-2:1234567890:cluster:my-cluster`

**Topics**
- Listing Tags Using the ElastiCache API (p. 437)
- Adding Tags Using the ElastiCache API (p. 438)
- Modifying Tags Using the ElastiCache API (p. 438)
- Removing Tags Using the ElastiCache API (p. 438)

### Listing Tags Using the ElastiCache API

You can use the ElastiCache API to list tags on an existing resource by using the `ListTagsForResource` operation.

The following code uses the ElastiCache API to list the tags on the resource `my-cluster-001` 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-001` in the us-west-2 region.

```
https://elasticache.us-west-2.amazonaws.com/
?Action=AddTagsToResource
&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 Redis node by using the `RemoveTagsFromResource` operation.

The following code uses the ElastiCache API to remove the tags with the keys `Service` and `Region` from the node `my-cluster-001` in the `my-cluster` cluster in region us-west-2.

```
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 upfront 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. 439)
- Legacy Reserved Node Cache Offerings (p. 439)
- Getting Info About Reserved Node Offerings (p. 442)
- Purchasing a Reserved Node (p. 445)
- Getting Info About Your Reserved Nodes (p. 448)

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 hourly fee. Applied to the whole term whether or not you're using the reserved node.</td>
<td>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 usage fee charged for each hour the node is running. No hourly charge when the node is not running.</td>
<td>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 usage fee charged for each hour the node is running. No hourly charge when the node is not running. Highest hourly fees of all the offering types, but fees apply only when the reserved node is running.</td>
<td>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 hourly fee. Applied whenever the node is running.</td>
<td>Highest hourly cost.</td>
</tr>
</tbody>
</table>
For more information, see Amazon ElastiCache Pricing.
Getting Info About Reserved Node Offerings

Getting Info About Reserved Node Offerings (Console)

To get pricing and other information about available reserved node 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.
4. For Product Description, choose Redis.
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": [
 { "RecurringChargeAmount": 0.491,
```

API Version 2015-02-02

442
For more information, see describe-reserved-cache-nodes-offerings in the AWS CLI Reference.

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:

  <DescribeReservedCacheNodesOfferingsResult>
    <ReservedCacheNodesOfferings>
      <ReservedCacheNodesOffering>
        <Duration>31536000</Duration>
        <OfferingType>Medium Utilization</OfferingType>
        <CurrencyCode>USD</CurrencyCode>
        <RecurringCharges/>
        <FixedPrice>1820.0</FixedPrice>
        <ProductDescription>memcached</ProductDescription>
      </ReservedCacheNodesOffering>
    </ReservedCacheNodesOfferings>
  </DescribeReservedCacheNodesOfferingsResult>
</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. 445)
- Purchasing a Reserved Node (AWS CLI) (p. 445)
- Purchasing a Reserved Node (ElastiCache API) (p. 446)

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:

```
aws elasticache purchase-reserved-cache-nodes-offering
   --reserved-cache-nodes-offering-id 649fd0c8-cf6d-47a0-bfa6-060f8e75e95f
   --reserved-cache-node-id myreservationID
```

For Windows:

```
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

446
<OfferingType>Medium Utilization</OfferingType>
<CurrencyCode>USD</CurrencyCode>
<RecurringCharges/>
<ProductDescription>memcached</ProductDescription>
<ReservedCacheNodesOfferingId>649fd0c8-cf6d-47a0-bfa6-060f8e75e95f</ReservedCacheNodesOfferingId>
<State>payment-pending</State>
<ReservedCacheNodeId>myreservationID</ReservedCacheNodeId>
<CacheNodeCount>10</CacheNodeCount>
<Duration>31536000</Duration>
<FixedPrice>123.0</FixedPrice>
<UsagePrice>0.123</UsagePrice>
<CacheNodeType>cache.m1.small</CacheNodeType>
</ReservedCacheNode>
</PurchaseReservedCacheNodesOfferingResult>
<ResponseMetadata>
<RequestId>7f099901-29cf-11e1-bd06-6fe008f046c3</RequestId>
</ResponseMetadata>
</PurchaseReservedCacheNodesOfferingResponse>

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. 448)
• Getting Info About Your Reserved Nodes (AWS CLI) (p. 448)
• Getting Info About Your Reserved Nodes (ElastiCache API) (p. 448)

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:

```bash
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

```bash
https://elasticache.us-west-2.amazonaws.com/
?Action=DescribeReservedCacheNodes
```

API Version 2015-02-02
448
This call returns output similar to the following:

```xml
  <DescribeReservedCacheNodesResult>
    <ReservedCacheNodes>
      <ReservedCacheNode>
        <OfferingType>Medium Utilization</OfferingType>
        <CurrencyCode>USD</CurrencyCode>
        <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>2340d50-2978-11e1-9ed-771388d6ed6b</RequestId>
  </ResponseMetadata>
</DescribeReservedCacheNodesResponse>
```

For more information, see `DescribeReservedCacheNodes` in the ElastiCache API Reference.

---

**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](https://aws.amazon.com/compliance/). For general information, see [AWS Compliance Programs](https://aws.amazon.com/compliance/).

You can download third-party audit reports using AWS Artifact. For more information, see [Downloading Reports in AWS Artifact](https://aws.amazon.com/artifact/download).

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.

**Topics**
- [ElastiCache for Redis Compliance](#) (p. 450)

---

### ElastiCache for Redis Compliance

In this section, you can find the compliance requirements and controls offered when using Amazon ElastiCache for Redis.

**Topics**
- [Self-Service Security Updates for Compliance](#) (p. 450)
- [ElastiCache for Redis FedRAMP Compliance](#) (p. 451)
- [HIPAA Eligibility](#) (p. 452)
- [ElastiCache for Redis PCI DSS Compliance](#) (p. 453)
- [Create and Seed a New Compliant Cluster](#) (p. 453)
- [More Information](#) (p. 453)

### Self-Service Security Updates for Compliance

ElastiCache offers a self-service software update feature called **Service Updates** via the Console, API and CLI. Using this feature, you can manage security updates on your Redis clusters on-demand and in real-time. This feature allows you to control when you update Redis clusters with the latest required security fixes, minimizing the impact on your business.

Security updates are released via the **Service Updates** feature. They are specified by the **Update Type** field of value **security update**. The Service Update has corresponding **Severity** and **Recommended Apply by Date** fields. In order to maintain compliance of your Redis clusters, you must apply the available updates by the **Recommended Apply by Date**. The field **SLA Met** reflects your Redis cluster’s compliance status.

**Note**

If you do not apply the Service Update by the recommended date or when the Service Update expires, ElastiCache will not take any action to apply the update on your behalf. You will be notified of the Service Updates applicable to your Redis clusters via an announcement on the Redis console, email, Amazon SNS, CloudWatch events and Personal Health Dashboard. For more information on Self-Service Maintenance see [Self-Service Updates in Amazon ElastiCache](#) (p. 474).
CloudWatch events and Personal Health Dashboard are not supported in the following regions:

- us-gov-west-1
- us-gov-east-1
- cn-north-1
- cn-northwest-1

**ElastiCache for Redis FedRAMP Compliance**

The AWS FedRAMP Compliance program includes Amazon ElastiCache for Redis as a FedRAMP-authorized service. If you are a federal or commercial customer, you can use the service to process and store sensitive workloads in AWS US East and US West with data up to the moderate impact level. You can use the service for sensitive workloads in the AWS GovCloud (US) Region's authorization boundary with data up to the high impact level.

You can request access to the AWS FedRAMP Security Packages through the FedRAMP PMO or your AWS Sales Account Manager or, they can be downloaded through **AWS Artifact** at [AWS Artifact](https://aws.amazon.com/artifact).

**Requirements**

To enable FedRAMP support on your ElastiCache for Redis cluster, your cluster and nodes within the cluster must satisfy the following requirements.

- **Engine version requirements** – Your cluster must be running ElastiCache for Redis 3.2.6, 4.0.10 and later for both cluster mode enabled and disabled to qualify for FedRAMP compliance.
  - Starting with ElastiCache for Redis versions 3.2.6, 4.0.10 and later for both cluster mode enabled and disabled, you can also enable additional data security features such as:
    - ElastiCache for Redis In-Transit Encryption (TLS) (p. 341)
    - At-Rest Encryption in ElastiCache for Redis (p. 348)
    - Authenticaing Users with the Redis AUTH Command (p. 354)
- **Node type requirements** – Your cluster must be running a current-generation node type — M4, M5, T2, R4 or R5. For more information, see the following:
  - Supported Node Types (p. 65)
  - Choosing Your Node Size (p. 76)
- **FIPS Endpoints requirements** – Your ElastiCache for Redis can be created using the FIPS endpoints available in the following regions:

<table>
<thead>
<tr>
<th>Region Name/Region</th>
<th>FIPS Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>US East (Ohio) Region</td>
<td>elasticache-fips.us-east-2.amazonaws.com</td>
</tr>
<tr>
<td>us-east-2</td>
<td></td>
</tr>
<tr>
<td>US East (N. Virginia) Region</td>
<td>elasticache-fips.us-east-1.amazonaws.com</td>
</tr>
<tr>
<td>us-east-1</td>
<td></td>
</tr>
<tr>
<td>US West (N. California) Region</td>
<td>elasticache-fips.us-west-1.amazonaws.com</td>
</tr>
<tr>
<td>us-west-1</td>
<td></td>
</tr>
<tr>
<td>US West (Oregon) Region</td>
<td>elasticache-fips.us-west-2.amazonaws.com</td>
</tr>
<tr>
<td>us-west-2</td>
<td></td>
</tr>
</tbody>
</table>
**HIPAA Eligibility**

The AWS HIPAA Compliance program includes Amazon ElastiCache for Redis as a HIPAA eligible service.

To use ElastiCache for Redis in compliance with HIPAA, you need to execute a Business Associate Agreement (BAA) with AWS. In addition, your cluster and the nodes within your cluster must satisfy the requirements for engine version, node type, and data security listed following.

**Requirements**

To enable HIPAA support on your ElastiCache for Redis cluster, your cluster and nodes within the cluster must satisfy the following requirements.

- **Engine version requirements** – Your cluster must be running one of the following ElastiCache for Redis versions to qualify for HIPAA eligibility.
  - ElastiCache for Redis Version 5.0.0 (Enhanced) (p. 48)
  - ElastiCache for Redis Version 4.0.10 (Enhanced) (p. 48)
  - ElastiCache for Redis Version 3.2.6 (Enhanced) (p. 49)

- **Node type requirements** – Your cluster must be running a current-generation node type—M4, M5, T2, R4 or R5. For more information, see the following:
  - Supported Node Types (p. 65)
  - Choosing Your Node Size (p. 76)

- **Data security requirements** – Your cluster must enable in-transit encryption, at-rest encryption, and Redis AUTH. For more information, see the following:
  - ElastiCache for Redis In-Transit Encryption (TLS) (p. 341)
  - At-Rest Encryption in ElastiCache for Redis (p. 348)
  - Authenticating Users with the Redis AUTH Command (p. 354)

- **Security Updates Requirement** – You must update your Redis cluster with the latest Service Updates of type security by the Recommended Apply by Date. You can update the cluster in real-time and on-demand to ensure no impact to your business. For more information, see Self-Service Updates in Amazon ElastiCache (p. 474)

By implementing these requirements, ElastiCache for Redis can be used to store, process, and access Protected Health Information (PHI) in compliance with HIPAA.

For general information about AWS Cloud and HIPAA eligibility, see the following:

- HIPAA Compliance
- Architecting for HIPAA Security and Compliance on Amazon Web Services
- Security Updates Requirement – You must regularly update your Redis cluster by the Recommended Apply by Date. You can update the cluster in real-time and on-demand to ensure no impact to your business. For more information, see Self-Service Updates in Amazon ElastiCache (p. 474).
ElastiCache for Redis PCI DSS Compliance

The AWS PCI DSS Compliance program includes Amazon ElastiCache for Redis as a PCI-compliant service. The PCI DSS 3.2 Compliance Package can be downloaded through AWS Artifact. For more information, see AWS PCI DSS Compliance Program.

Requirements

To enable PCI DSS support on your ElastiCache for Redis cluster, your cluster and nodes within the cluster must satisfy the following requirements.

- **Engine version requirements** – Your cluster must be running ElastiCache for Redis 3.2.6, 4.0.10 and later for both cluster mode enabled and disabled.

- **Node type requirements** – Your cluster must be running a current-generation node type—M4, M5, T2, R4 or R5. For more information, see the following:
  - Supported Node Types (p. 65)
  - Choosing Your Node Size (p. 76)

- **Security Updates Requirement** – You must regularly update your Redis cluster by the Recommended Apply by Date. You can update the cluster in real-time and on-demand to ensure no impact to your business. For more information, see Self-Service Updates in Amazon ElastiCache (p. 474).

ElastiCache for Redis also offers Data Security Controls to further secure the cluster to store, process, and transmit sensitive financial data like Customer Cardholder Data (CHD) when using the service.

**Data security options** – For more information, see the following:

- ElastiCache for Redis In-Transit Encryption (TLS) (p. 341)
- At-Rest Encryption in ElastiCache for Redis (p. 348)
- Authenticating Users with the Redis AUTH Command (p. 354)

Create and Seed a New Compliant Cluster

To create a compliant cluster, create a new cluster and make sure that your choices fulfill the requirements for the compliance you want. These requirements can include engine version, node type, encryption, and if needed FIPS endpoints. If you choose, you can seed a new compliant cluster with data from an existing cluster as you’re creating it. For more information, see the following:

- Creating a Cluster (p. 74)
- Creating a Redis Replication Group from Scratch (p. 168)
- Seeding a New Cluster with an Externally Created Backup (p. 242)

More Information

For general information about AWS Cloud compliance, see the following:

- Self-Service Security Updates for Compliance (p. 450)
- Self-Service Updates in Amazon ElastiCache (p. 474)
- AWS Cloud Compliance
- Shared Responsibility Model
- AWS Services in Scope by Compliance Program
- AWS HIPAA Compliance Program
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. 454)

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 Redis (p. 454)
- Recommendations (p. 456)

Mitigating Failures when Running Redis

When running the Redis engine, you have the following options for minimizing the impact of a node or Availability Zone failure.

Mitigating Node Failures

To mitigate the impact of Redis node failures, you have the following options:

Topics
- Mitigating Failures: Redis Append Only Files (AOF) (p. 454)
- Mitigating Failures: Redis Replication Groups (p. 455)

Mitigating Failures: Redis Append Only Files (AOF)

When AOF is enabled for Redis, whenever data is written to your Redis cluster, a corresponding transaction record is written to a Redis append only file (AOF). If your Redis process restarts, ElastiCache creates a replacement cluster and provisions it. You can then run the AOF against the cluster to repopulate it with data.

Some of the shortcomings of using Redis AOF to mitigate cluster failures are the following:
• It is time-consuming.

Creating and provisioning a cluster can take several minutes. Depending on the size of the AOF, running it against the cluster adds even more time when your application can't access your cluster for data. This forces your application to hit the database directly.

• The AOF can get big.

Because every write to your cluster is written to a transaction record, AOFs can become very large, larger than the .rdb file for the dataset in question. Because ElastiCache relies on the local instance store, which is limited in size, enabling AOF can cause out-of-disk-space issues. You can avoid out-of-disk-space issues by using a replication group with Multi-AZ enabled.

• Using AOF can't protect you from all failure scenarios.

For example, if a node fails due to a hardware fault in an underlying physical server, ElastiCache will provision a new node on a different server. In this case, the AOF is not available and can't be used to recover the data.

For more information, see Append Only Files (AOF) in ElastiCache for Redis (p. 250).

Mitigating Failures: Redis Replication Groups

A Redis replication group is comprised of a single primary node which your application can both read from and write to, and from 1 to 5 read-only replica nodes. Whenever data is written to the primary node it is also asynchronously updated on the read replica nodes.

When a read replica fails

1. ElastiCache detects the failed read replica.
2. ElastiCache takes the failed node off line.
3. ElastiCache launches and provisions a replacement node in the same AZ.
4. The new node synchronizes with the Primary node.

During this time your application can continue reading and writing using the other nodes.

Redis Multi-AZ with Automatic Failover

You can enable Multi-AZ with automatic failover on your Redis replication groups. Whether you enable Multi-AZ with auto failover or not, a failed Primary will be detected and replaced automatically. How this takes place varies whether or not Multi-AZ is or is not enabled.

When Multi-AZ with Auto Failover is enabled

1. ElastiCache detects the Primary node failure.
2. ElastiCache promotes the read replica node with the least replication lag to primary node.
3. The other replicas sync with the new primary node.
4. ElastiCache spins up a read replica in the failed primary's AZ.
5. The new node syncs with the newly promoted primary.

Failing over to a replica node is generally faster than creating and provisioning a new Primary node. This means your application can resume writing to your Primary node sooner than if Multi-AZ were not enabled.
For more information, see Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148).

**When Multi-AZ with Auto Failover is disabled**

1. ElastiCache detects Primary failure.
2. ElastiCache takes the Primary offline.
3. ElastiCache creates and provisions a new Primary node to replace the failed Primary.
4. ElastiCache syncs the new Primary with one of the existing replicas.
5. When the sync is finished, the new node functions as the cluster's Primary node.

During this process, steps 1 through 4, your application can't write to the Primary node. However, your application can continue reading from your replica nodes.

For added protection, we recommend that you launch the nodes in your replication group in different Availability Zones (AZs). If you do this, an AZ failure will only impact the nodes in that AZ and not the others.

For more information, see High Availability Using Replication Groups (p. 142).

**Mitigating Availability Zone Failures**

To mitigate the impact of an Availability Zone failure, locate your nodes in as many Availability Zones as possible.

No matter how many nodes you have, if they are all located in the same Availability Zone, a catastrophic failure of that AZ results in your losing all your cache data. However, if you locate your nodes in multiple AZs, a failure of any AZ results in your losing only the nodes in that AZ.

Any time you lose a node you can experience a performance degradation since read operations are now shared by fewer nodes. This performance degradation will continue until the nodes are replaced. Because your data is not partitioned across Redis nodes, you risk some data loss only when the primary node is lost.

For information on specifying the Availability Zones for Redis nodes, see Creating a Redis (cluster mode disabled) Cluster (Console) (p. 79).

For more information on regions and Availability Zones, see Choosing Regions and Availability Zones (p. 55).

**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 Redis, we recommend that you enable Multi-AZ on your replication group so that ElastiCache will automatically fail over to a replica if the primary 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.
Other precautions

If you're running Redis, then in addition to the above, we recommend that you schedule regular backups of your cluster. Backups (snapshots) create a .rdb file you can use to restore your cluster in case of failure or corruption. For more information, see Backup and Restore for ElastiCache for Redis (p. 213).

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.
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. 458)
- Restricted Redis Commands (p. 463)
- Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464)
- Managing Reserved Memory (p. 466)
- Mitigating Out-of-Disk-Space Issues When Using Redis AOF (p. 471)
- Best Practices: Online Cluster Resizing (p. 472)
- Best Practices: Minimizing Downtime During Maintenance (p. 473)

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. 458)
- Write-Through (p. 460)
- Adding TTL (p. 461)
- Related Topics (p. 462)

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.

![Diagram of cache operations](image)

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. 460) and *Adding TTL* (p. 461) strategies.
Lazy Loading Pseudocode Example

The following is a pseudocode example of lazy loading logic.

```plaintext
// *****************************************
// 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
// *****************************************
get_customer(customer_id)

    customer_record = cache.get(customer_id)
    if (customer_record == null)
        customer_record = db.query("SELECT * FROM Customers WHERE id == {0}", customer_id)
        cache.set(customer_id, customer_record)
    return customer_record
```

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

```plaintext
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. 458) 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. 461), you can minimize wasted space.

**Write-Through Pseudocode Example**

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

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

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

```c
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 Redis `set` command.

**TTL Pseudocode Examples**

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

```c
// *****************************************
// 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}", customer_id, values)
    cache.set(customer_id, customer_record, 300)
    return success
```

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

```c
// *****************************************
```
// 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, 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.

```java
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 Redis Clusters (p. 250)
Restricted Redis Commands

To deliver a managed service experience, ElastiCache restricts access to certain cache engine-specific commands that require advanced privileges. For cache clusters running Redis, the following commands are unavailable:

- bgrewriteaof
- bgsave
- cluster addslot
- cluster delslot
- cluster setslot
- config
- debug
- migrate
- replicaof
- save
- slaveof
- shutdown
- sync
Ensuring That You Have Enough Memory to Create a Redis Snapshot

Redis snapshots and synchronizations in version 2.8.22 and later

Redis 2.8.22 introduces a forkless save process that allows you to allocate more of your memory to your application's use without incurring increased swap usage during synchronizations and saves. For more information, see How Synchronization and Backup are Implemented (p. 161).

Redis snapshots and synchronizations before version 2.8.22

When you work with Redis ElastiCache, Redis calls a background write command in a number of cases:

- When creating a snapshot for a backup.
- When synchronizing replicas with the primary in a replication group.
- When enabling the append-only file feature (AOF) for Redis.
- When promoting a replica to master (which causes a primary/replica sync).

Whenever Redis executes a background write process, you must have sufficient available memory to accommodate the process overhead. Failure to have sufficient memory available causes the process to fail. Because of this, it is important to choose a node instance type that has sufficient memory when creating your Redis cluster.

Background Write Process and Memory Usage

Whenever a background write process is called, Redis forks its process (remember, Redis is single threaded). One fork persists your data to disk in a Redis .rdb snapshot file. The other fork services all read and write operations. To ensure that your snapshot is a point-in-time snapshot, all data updates and additions are written to an area of available memory separate from the data area.

As long as you have sufficient memory available to record all write operations while the data is being persisted to disk, you should have no insufficient memory issues. You are likely to experience insufficient memory issues if any of the following are true:

- Your application performs many write operations, thus requiring a large amount of available memory to accept the new or updated data.
- You have very little memory available in which to write new or updated data.
- You have a large dataset that takes a long time to persist to disk, thus requiring a large number of write operations.

The following diagram illustrates memory use when executing a background write process.
Avoiding Running Out of Memory When Executing a Background Write

Whenever a background write process such as BGSAVE or BGREWRITEAOF is called, to keep the process from failing, you must have more memory available than will be consumed by write operations during the process. The worst-case scenario is that during the background write operation every Redis record is updated and some new records are added to the cache. Because of this, we recommend that you set reserved-memory-percent to 50 (50 percent) for Redis versions before 2.8.22 or 25 (25 percent) for Redis versions 2.8.22 and later.

The maxmemory value indicates the memory available to you for data and operational overhead. Because you cannot modify the reserved-memory parameter in the default parameter group, you must create a custom parameter group for the cluster. The default value for reserved-memory is 0, which allows Redis to consume all of maxmemory with data, potentially leaving too little memory for other uses, such as a background write process. For maxmemory values by node instance type, see Redis Node-Type Specific Parameters (p. 336).

You can also use reserved-memory parameter to reduce the amount of memory Redis uses on the box.

For more information on Redis-specific parameters in ElastiCache, see Redis Specific Parameters (p. 315).

For information on creating and modifying parameter groups, see Creating a Parameter Group (p. 300) and Modifying a Parameter Group (p. 310).
Managing Reserved Memory

Reserved memory is memory set aside for nondata use. When performing a backup or failover, Redis uses available memory to record write operations to your cluster while the cluster's data is being written to the .rdb file. If you don't have sufficient memory available for all the writes, the process fails. Following, you can find information on options for managing reserved memory for ElastiCache for Redis and how to apply those options.

Topics
- How Much Reserved Memory Do You Need? (p. 466)
- Parameters to Manage Reserved Memory (p. 466)
- Specifying Your Reserved Memory Management Parameter (p. 469)

How Much Reserved Memory Do You Need?

If you are running a version of Redis before 2.8.22, reserve more memory for backups and failovers than if you are running Redis 2.8.22 or later. This requirement is due to the different ways that ElastiCache for Redis implements the backup process. The rule of thumb is to reserve half of a node type's `maxmemory` value for Redis overhead for versions before 2.8.22, and one-fourth for Redis versions 2.8.22 and later.

For more information, see the following:
- Ensuring That You Have Enough Memory to Create a Redis Snapshot (p. 464)
- How Synchronization and Backup are Implemented (p. 161)

Parameters to Manage Reserved Memory

As of March 16, 2017, Amazon ElastiCache for Redis provides two mutually exclusive parameters for managing your Redis memory, `reserved-memory` and `reserved-memory-percent`. Neither of these parameters is part of the Redis distribution.

Depending upon when you became an ElastiCache customer, one or the other of these parameters is the default memory management parameter. This parameter applies when you create a new Redis cluster or replication group and use a default parameter group.

- For customers who started before March 16, 2017 – When you create a Redis cluster or replication group using the default parameter group, your memory management parameter is `reserved-memory`. In this case, zero (0) bytes of memory are reserved.
- For customers who started on or after March 16, 2017 – When you create a Redis cluster or replication group using the default parameter group, your memory management parameter is `reserved-memory-percent`. In this case, 25 percent of your node's `maxmemory` value is reserved for nondata purposes.

After reading about the two Redis memory management parameters, you might prefer to use the one that isn't your default or with nondefault values. If so, you can change to the other reserved memory management parameter.

To change the value of that parameter, you can create a custom parameter group and modify it to use your preferred memory management parameter and value. You can then use the custom parameter group whenever you create a new Redis cluster or replication group. For existing clusters or replication groups, you can modify them to use your custom parameter group.
For more information, see the following:

- Specifying Your Reserved Memory Management Parameter (p. 469)
- Creating a Parameter Group (p. 300)
- Modifying a Parameter Group (p. 310)
- Modifying an ElastiCache Cluster (p. 100)
- Modifying a Replication Group (p. 188)

The reserved-memory Parameter

Before March 16, 2017, all ElastiCache for Redis reserved memory management was done using the parameter reserved-memory. The default value of reserved-memory is 0. This default reserves no memory for Redis overhead and allows Redis to consume all of a node's memory with data.

Changing reserved-memory so you have sufficient memory available for backups and failovers requires you to create a custom parameter group. In this custom parameter group, you set reserved-memory to a value appropriate for the Redis version running on your cluster and cluster's node type. For more information, see How Much Reserved Memory Do You Need? (p. 466)

The ElastiCache for Redis parameter reserved-memory is specific to ElastiCache for Redis and isn't part of the Redis distribution.

The following procedure shows how to use reserved-memory to manage the memory on your Redis cluster.

To reserve memory using reserved-memory

1. Create a custom parameter group specifying the parameter group family matching the engine version you're running—for example, specifying the redis2.8 parameter group family. For more information, see Creating a Parameter Group (p. 300).

   ```bash
   aws elasticache create-cache-parameter-group \
   --cache-parameter-group-name redis28-m3xl \
   --description "Redis 2.8.x for m3.xlarge node type" \
   --cache-parameter-group-family redis2.8
   ```

2. Calculate how many bytes of memory to reserve for Redis overhead. You can find the value of maxmemory for your node type at Redis Node-Type Specific Parameters (p. 336).

3. Modify the custom parameter group so that the parameter reserved-memory is the number of bytes you calculated in the previous step. The following AWS CLI example assumes you're running a version of Redis before 2.8.22 and need to reserve half of the node's maxmemory. For more information, see Modifying a Parameter Group (p. 310).

   ```bash
   aws elasticache modify-cache-parameter-group \
   --cache-parameter-group-name redis28-m3xl \
   --parameter-name-values "ParameterName=reserved-memory, ParameterValue=7130316800"
   ```

   You need a separate custom parameter group for each node type that you use, because each node type has a different maxmemory value. Thus, each node type needs a different value for reserved-memory.

4. Modify your Redis cluster or replication group to use your custom parameter group.

   The following CLI example modifies the cluster my-redis-cluster to use the custom parameter group redis28-m3xl beginning immediately. For more information, see Modifying an ElastiCache Cluster (p. 100).
The reserved-memory-percent parameter

On March 16, 2017, Amazon ElastiCache introduced the parameter reserved-memory-percent and made it available on all versions of ElastiCache for Redis. The purpose of reserved-memory-percent is to simplify reserved memory management across all your clusters. It does so by enabling you to have a single parameter group for each parameter group family (such as redis2.8) to manage your clusters' reserved memory, regardless of node type. The default value for reserved-memory-percent is 25 (25 percent).

The ElastiCache for Redis parameter reserved-memory-percent is specific to ElastiCache for Redis and isn't part of the Redis distribution.

To reserve memory using reserved-memory-percent

To use reserved-memory-percent to manage the memory on your ElastiCache for Redis cluster, do one of the following:

- If you are running Redis 2.8.22 or later, assign the default parameter group to your cluster. The default 25 percent should be adequate. If not, take the steps described following to change the value.
- If you are running a version of Redis before 2.8.22, you probably need to reserve more memory than reserved-memory-percent's default 25 percent. To do so, use the following procedure.

To change the percent value of reserved-memory-percent

1. Create a custom parameter group specifying the parameter group family matching the engine version you're running—for example, specifying the redis2.8 parameter group family. A custom parameter group is necessary because you can't modify a default parameter group. For more information, see Creating a Parameter Group (p. 300).

```bash
aws elasticache create-cache-parameter-group \
  --cache-parameter-group-name redis28-50 \
  --description "Redis 2.8.x 50% reserved" \
  --cache-parameter-group-family redis2.8
```

Because reserved-memory-percent reserves memory as a percent of a node's maxmemory, you don't need a custom parameter group for each node type.

2. Modify the custom parameter group so that reserved-memory-percent is 50 (50 percent). For more information, see Modifying a Parameter Group (p. 310).

```bash
aws elasticache modify-cache-parameter-group \
```
Specifying Your Reserved Memory Management Parameter

If you were a current ElastiCache customer on March 16, 2017, your default reserved memory management parameter is reserved-memory with zero (0) bytes of reserved memory. If you became an ElastiCache customer after March 16, 2017, your default reserved memory management parameter is reserved-memory-percent with 25 percent of the node's memory reserved. This is true no matter when you created your ElastiCache for Redis cluster or replication group. However, you can change your reserved memory management parameter using either the AWS CLI or ElastiCache API.

The parameters reserved-memory and reserved-memory-percent are mutually exclusive. A parameter group always has one but never both. You can change which parameter a parameter group uses for reserved memory management by modifying the parameter group. The parameter group must be a custom parameter group, because you can't modify default parameter groups. For more information, see Creating a Parameter Group (p. 300).

To specify reserved-memory-percent

To use reserved-memory-percent as your reserved memory management parameter, modify a custom parameter group using the modify-cache-parameter-group command. Use the parameter-name-values parameter to specify reserved-memory-percent and a value for it.

```
aws elasticache modify-cache-parameter-group \
   --cache-parameter-group-name redis32-cluster-on \
   --parameter-name-values "ParameterName=reserved-memory-percent, ParameterValue=25"
```

To specify reserved-memory

3. Use this custom parameter group for any Redis clusters or replication groups running a version of Redis older than 2.8.22.

The following CLI example modifies the Redis cluster my-redis-cluster to use the custom parameter group redis28-50 beginning immediately. For more information, see Modifying an ElastiCache Cluster (p. 100).

```
aws elasticache modify-cache-cluster \
   --cache-cluster-id my-redis-cluster \
   --cache-parameter-group-name redis28-50 \
   --apply-immediately
```

The following CLI example modifies the Redis replication group my-redis-repl-grp to use the custom parameter group redis28-50 beginning immediately. For more information, see Modifying a Replication Group (p. 188).

```
aws elasticache modify-replication-group \
   --replication-group-id my-redis-repl-grp \
   --cache-parameter-group-name redis28-50 \
   --apply-immediately
```
To use reserved-memory as your reserved memory management parameter, modify a custom parameter group using the `modify-cache-parameter-group` command. Use the `parameter-name-values` parameter to specify `reserved-memory` and a value for it.

The following CLI example modifies the custom parameter group `redis32-m3x1` so that it uses `reserved-memory` to manage reserved memory. A value must be assigned to `ParameterValue` for the parameter group to use the `ParameterName` parameter for reserved memory management. Because the engine version is newer than 2.8.22, we set the value to 3565158400 which is 25 percent of a `cache.m3.xlarge`'s `maxmemory`. For more information, see Modifying a Parameter Group (p. 310).

```bash
aws elasticache modify-cache-parameter-group \
  --cache-parameter-group-name redis32-m3x1 \
  --parameter-name-values "ParameterName=reserved-memory, ParameterValue=3565158400"
```
Mitigating Out-of-Disk-Space Issues When Using Redis AOF

When planning your Amazon ElastiCache implementation, you should plan so that failures have the least impact possible.

You enable AOF because an AOF file is useful in recovery scenarios. In case of a node restart or service crash, Redis replays the updates from an AOF file, thereby recovering the data lost due to the restart or crash.

**Warning**

AOF cannot protect against all failure scenarios. For example, if a node fails due to a hardware fault in an underlying physical server, ElastiCache provisions a new node on a different server. In this case, the AOF file is no longer available and cannot be used to recover the data. Thus, Redis restarts with a cold cache.

Enabling Redis Multi-AZ as a Better Approach to Fault Tolerance

If you are enabling AOF to protect against data loss, consider using a replication group with Multi-AZ enabled instead of AOF. When using a Redis replication group, if a replica fails, it is automatically replaced and synchronized with the primary cluster. If Multi-AZ is enabled on a Redis replication group and the primary fails, it fails over to a read replica. Generally, this functionality is much faster than rebuilding the primary from an AOF file. For greater reliability and faster recovery, we recommend that you create a replication group with one or more read replicas in different Availability Zones and enable Multi-AZ instead of using AOF. Because there is no need for AOF in this scenario, ElastiCache disables AOF on Multi-AZ replication groups.

For more information, see the following topics:

- [Mitigating Failures](#) (p. 454)
- [High Availability Using Replication Groups](#) (p. 142)
- [Minimizing Downtime: Multi-AZ with Automatic Failover](#) (p. 148)
Best Practices: Online Cluster Resizing

Resharding involves adding and removing shards or nodes to your cluster and redistributing key spaces. As a result, multiple things have an impact on the resharding operation, such as the load on the cluster, memory utilization, and overall size of data. For the best experience, we recommend that you follow overall cluster best practices for uniform workload pattern distribution. In addition, we recommend taking the following steps.

Before initiating resharding, we recommend the following:

- **Test your application** – Test your application behavior during resharding in a staging environment if possible.

- **Get early notification for scaling issues** – Resharding is a compute-intensive operation. Because of this, we recommend keeping CPU utilization under 80 percent on multicore instances and less than 50 percent on single core instances during resharding. Monitor ElastiCache for Redis metrics and initiate resharding before your application starts observing scaling issues. Useful metrics to track are CPUUtilization, NetworkBytesIn, NetworkBytesOut, CurrConnections, NewConnections, FreeableMemory, SwapUsage, and BytesUsedForCache.

- **Ensure sufficient free memory is available before scaling in** – If you're scaling in, ensure that free memory available on the shards to be retained is at least 1.5 times the memory used on the shards you plan to remove.

- **Initiate resharding during off-peak hours** – This practice helps to reduce the latency and throughput impact on the client during the resharding operation. It also helps to complete resharding faster as more resources can be used for slot redistribution.

- **Review client timeout behavior** – Some clients might observe higher latency during online cluster resizing. Configuring your client library with a higher timeout can help by giving the system time to connect even under higher load conditions on server. In some cases, you might open a large number of connections to the server. In these cases, consider adding exponential backoff to reconnect logic. Doing this can help prevent a burst of new connections hitting the server at the same time.

During resharding, we recommend the following:

- **Avoid expensive commands** – Avoid running any computationally and I/O intensive operations, such as the KEYS and SMEMBERS commands. We suggest this approach because these operations increase the load on the cluster and have an impact on the performance of the cluster. Instead, use the SCAN and SSCAN commands.

- **Follow Lua best practices** – Avoid long running Lua scripts, and always declare keys used in Lua scripts up front. We recommend this approach to determine that the Lua script is not using cross slot commands. Ensure that the keys used in Lua scripts belong to the same slot.

After resharding, note the following:

- Scale-in might be partially successful if insufficient memory is available on target shards. If such a result occurs, review available memory and retry the operation, if necessary.

- Slots with large items are not migrated. In particular, slots with items larger than 256 MB post-serialization are not migrated.

- The BRPOPPLPUSH command is not supported if it operates on the slot being migrated. FLUSHALL and FLUSHDB commands are not supported inside Lua scripts during a resharding operation.
Best Practices: Minimizing Downtime During Maintenance

Cluster mode configuration has the best availability during managed or unmanaged operations. We recommend that you use a cluster mode supported client that connects to the cluster discovery endpoint. For cluster mode disabled, we recommend that you use the primary endpoint for all write operations.

For read activity, applications can also connect to any node in the cluster. Unlike the primary endpoint, node endpoints resolve to specific endpoints. If you make a change in your cluster, such as adding or deleting a replica, you must update the node endpoints in your application.

If autofailover is enabled in the cluster, the primary node might change. Therefore, the application should confirm the role of the node and update all the read endpoints. Doing this helps ensure that you aren't causing a major load on the primary. With autofailover disabled, the role of the node doesn't change. However, the downtime in managed or unmanaged operations is higher as compared to clusters with auto failover enabled.

Avoid directing read requests to read replicas only. If you configure your client to direct read requests to read replicas only, ensure that you have at least two read replicas to avoid any read interruption during maintenance.
Self-Service Updates in Amazon ElastiCache

Amazon ElastiCache automatically monitors your fleet of Redis 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 Redis clusters in real-time. Depending on your business requirements, you can choose to stop the update to remaining nodes and 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 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. This dashboard view enables you to view all the service updates and their status for to your ElastiCache Redis fleet. It’s an audit log that you can use when reviewing your Redis fleet for service updates.

**Note**
Using this log can prove important when reviewing your fleet for compliance. For more information, see Self-Service Security Updates for Compliance (p. 450).

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 Redis clusters are always up-to-date with current security patches. To view the up-to-date status of all your Redis clusters, you can choose Service Update Status. This view also shows the clusters for which the update is not applicable. In addition, you might find updates that you applied to Redis clusters exceed their estimated update time and interrupt your business flows. In this case, you can stop them and reapply them at a time that better suits your business needs.

For more information, see Amazon ElastiCache Maintenance Help Page.

The following sections explore these options in detail.

**Topics**
- Applying the Self-Service Updates (p. 474)
- Stopping the Self-Service Updates (p. 480)

**Applying the Self-Service Updates**

You can start applying the service updates to your Redis 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 Redis fleet and applying any service-specific updates to applicable Redis clusters, see Applying the Service Updates Using the Console for Redis (p. 475).

When a new service update is available for one or more Redis 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.

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 Redis (p. 475)
  • Applying the Service Updates Using the Service Updates List (p. 477)

Applying the Service Updates Using the Console for Redis

Choose this to review the Update Status of individual Redis 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 Redis 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 Redis cluster and the **Estimated Update Time** enables you to plan your maintenance schedule. If service updates exceed the estimated time constraints for your business flows, you can stop them and reapply them at a later date. For more information, see **Stopping the Self-Service Updates** (p. 480).

- If you choose to apply the service updates to any or all available Redis 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.
On the ElastiCache dashboard, you can check **Update Status** for each of your Redis clusters. Using **Update Status**, you can find out the compliance status of your clusters for 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 the duration of the Estimated Update Time.
- **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 Redis 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 compliant. For more information about compliance, see **Self-Service Security Updates for Compliance (p. 450)**.

**Applying the Service Updates Using the Service Updates List**

To see the list of individual service updates and their status, along with other relevant information, choose the **Service Updates List** tab.
The Service Updates List contains the following:

- **Service Update Name**: A unique identifier for the service update.
- **Status**: The status of the update, which is one of the following:
  - **available**: The update is available for requisite Redis clusters.
  - **complete**: The update has been applied and all Redis clusters are compliant. (For more information, see Self-Service Security Updates for Compliance (p. 450)).
  - **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.
Important
We strongly recommend that you apply updates of type security as soon as your business flows allow. Doing this ensures that your Redis clusters are always up-to-date with the latest security patches and are compliant. For more information, see Self-Service Security Updates for Compliance (p. 450).

To review the list of individual service updates in relation to the applicable Redis 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 Redis 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 Redis clusters.
  - complete: The update has been applied and all Redis clusters are Compliant.
  - 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 by the apply-by date.
  - 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.
• **N/A**: The replication group was created after the recommended apply-by date.

• **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:

```bash
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:

```bash
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:

```bash
aws elasticache batch-apply-update-action --service-update-name sample-service-update --replication-group-ids my-replication-group-1 my-replication-group-2
```

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 Redis (p. 475)](#).

### Stopping the Service Updates Using the Console

You can interrupt a service update using the Redis console. The following demonstrates how to do this:

• After a service update has progressed on a selected Redis cluster, the ElastiCache console displays the **View/Stop Update** tab at the top of the Redis dashboard.
To interrupt the update, choose **Stop Update**.

When you stop the update, choose the Redis cluster and examine the status. It reverts to a **Stopping** status, as shown following, 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 --replication-group-ids my-replication-group-1 my-replication-group-2
```

For more information, see BatchStopUpdateAction.
Reference

The topics in this section cover working with the Amazon ElastiCache API and the ElastiCache section of the AWS CLI. Also included in this section are common error messages and service notifications.

- Using the ElastiCache API (p. 483)
- ElastiCache API Reference
- ElastiCache section of the AWS CLI Reference
- Amazon ElastiCache Error Messages (p. 493)
- Notifications (p. 494)

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. 483)
- Available Libraries (p. 485)
- Troubleshooting Applications (p. 486)
- Logging Amazon ElastiCache API Calls with AWS CloudTrail (p. 486)

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).

```plaintext
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).

```plaintext
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/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
```

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=2877960fced9040b41b4feaca835fd5cfe9264f768e6a0236c9145f915ffa56
```

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**.
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.

```perl
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...
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.

```
{
  "eventVersion": "1.01",
  "userIdentity": {
    "type": "IAMUser",
    "principalId": "EXAMPLEEXAMPLEEXAMPLE",
    "arn": "arn:aws:iam::123456789012:user/elasticache-allow",
    "accountId": "123456789012",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
    "userName": "elasticache-allow"
  },
  "eventTime": "2014-12-01T01:00:00Z",
  "eventSource": "elasticache.amazonaws.com",
  "eventName": "DescribeCacheClusters",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "192.0.2.01",
  "userAgent": "Amazon CLI/ElastiCache 1.10 API 2014-12-01",
  "requestParameters": {
    "cacheClusterId": "test-memcached",
    "engine": "memcached",
    "aZMode": "cross-az",
    "cacheNodeType": "cache.m1.small"
  },
  "responseElements": {
    "engine": "memcached",
    "clientDownloadLandingPage": "https://console.aws.amazon.com/elasticache/home#client-download",
    "cacheParameterGroup": {
      "cacheParameterGroupName": "default.memcached1.4",
      "cacheNodeIdsToReboot": [],
      "parameterApplyStatus": "in-sync"
    },
    "preferredAvailabilityZone": "Multiple",
    "numCacheNodes": 2,
    "cacheNodeType": "cache.m1.small",
    "cacheClusterStatus": "creating",
    "autoMinorVersionUpgrade": true,
    "preferredMaintenanceWindow": "thu:05:00-thu:06:00",
    "cacheClusterId": "test-memcached",
    "engineVersion": "1.4.14",
    "cacheSecurityGroups": [
      {
        "status": "active",
        "cacheSecurityGroupName": "default"
      }
    ],
    "pendingModifiedValues": {}
  },
  "requestID": "104f30b3-3548-11e4-b7b8-6d79ffe84edd",
  "eventID": "92762127-7a68-42ce-8787-927d2174cde1"
}
```
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",
      "cacheNodeIdToReboot": {}
    },
    "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-Thur: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": {
```
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. 490)
- Getting the Command Line Tools (p. 491)
- Setting Up the Tools (p. 491)
- Providing Credentials for the Tools (p. 492)
- Environmental Variables (p. 493)

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
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:
    
    ```bash
    $ export PATH=$PATH:$AWS_ELASTICACHE_HOME/bin
    ```
  
  - On Windows, enter the following command:
    
    ```cmd
    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:
     
     ```bash
     $ export AWS_CREDENTIAL_FILE=<the file created above>
     ```
   
   - On Windows, set the variable using the following command:
     
     ```cmd
     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 Redis. 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. 493)
- **Customer’s node quota exceeded** (p. 494)
- **Manual snapshot quota exceeded** (p. 494)
- **Insufficient cache cluster capacity** (p. 494)

**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 \%n 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: **The maximum number of manual snapshots for this cluster taken within 24 hours has been reached** or **The maximum number of manual snapshots for this node taken within 24 hours has been reached its quota of \%n**

**Cause:** You attempted to take a manual snapshot of a cluster when you have already taken the maximum number of manual snapshots allowed in a 24-hour period.

**Solution:** Wait 24 hours to attempt another manual snapshot of the cluster. Or, if you need to take a manual snapshot now, take the snapshot of another node that has the same data, such as a different node in a cluster.

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 Redis Clusters (p. 250).

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 Redis Specific Notifications

There are currently no outstanding ElastiCache for Redis notifications.
ElastiCache for Redis Documentation History

- **API version**: 2015-02-02
- **Latest documentation update**: March 16, 2020

The following table describes important changes in each release of the *ElastiCache for Redis User Guide* after March 2018. For notification about updates to this documentation, you can subscribe to the RSS feed.

**Recent ElastiCache for Redis Updates**

<table>
<thead>
<tr>
<th>update-history-change</th>
<th>update-history-description</th>
<th>update-history-date</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElastiCache for Redis now supports Global Datastore for Redis (p. 496)</td>
<td>The Global Datastore for Redis feature offers 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. You can create, modify, and describe a global datastore. You can also add or remove AWS Regions from your global datastore and promote an AWS Region as primary within a global datastore. For more information, see Replication Across AWS Regions Using Global Datastore.</td>
<td>March 16, 2020</td>
</tr>
<tr>
<td>ElastiCache for Redis now supports Redis version 5.0.6 (p. 496)</td>
<td>For more information, see ElastiCache for Redis Version 5.0.6 (Enhanced).</td>
<td>December 18, 2019</td>
</tr>
<tr>
<td>Amazon ElastiCache now supports T3-Standard cache nodes (p. 496)</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>
</tbody>
</table>
Amazon ElastiCache now supports modifying the AUTH token on an existing ElastiCache for Redis server (p. 496)

ElastiCache for Redis 5.0.5 now enables you to modify authentication tokens by setting and rotating new tokens. You can now modify active tokens while they're in use. You can also add brand-new tokens to existing clusters enabled with encryption in transit that were previously set up without authentication tokens. This is a two-step process by which you can set and rotate the token without interrupting client requests. This feature is currently not supported on AWS CloudFormation. For more information, see Authenticating Users with the Redis AUTH Command.

Amazon ElastiCache now supports online data migration from Redis on Amazon EC2 (p. 496)

You can now use Online Migration to migrate your data from self-hosted Redis on Amazon EC2 to Amazon ElastiCache. For more information, see Online Migration to ElastiCache.

ElastiCache for Redis now supports Redis version 5.0.5 (p. 496)

This update includes online configuration changes for ElastiCache for Redis of autofailover clusters during all planned operations. For more information, see ElastiCache for Redis Version 5.0.5 (Enhanced).

ElastiCache for Redis introduces online vertical scaling for Redis Cluster mode. (p. 496)

You can now scale up or scale down your sharded Redis Cluster on demand. ElastiCache for Redis resizes your cluster by changing the node type, while the cluster continues to stay online and serve incoming requests. For more information, see Online Vertical Scaling by Modifying Node Type.

ElastiCache for Redis now allows users to use a single reader endpoint for your Amazon ElastiCache for Redis cluster. (p. 496)

This feature allows you to direct all read traffic to your ElastiCache for Redis cluster through a single, cluster-level endpoint to take advantage of load balancing and higher availability. For more information, see Finding Connection Endpoints.

October 30, 2019

October 28, 2019

September 23, 2019

August 20, 2019

June 13, 2019
ElastiCache for Redis now allows users to apply service updates on their own schedule (p. 496) 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 if your cluster is in ElastiCache-supported compliance programs. For more information, see Self-Service Updates in Amazon ElastiCache and Self-Service Security Updates for Compliance. June 4, 2019

ElastiCache for Redis now supports Redis version 5.0.4, including engine stability guarantee in special conditions. (p. 496) This also includes improved Hyperloglog error handling and other enhancements. For more information, see Redis 5.0.4 release notes. May 15, 2019

ElastiCache for Redis now supports Redis version 5.0.3 and ability to rename commands. (p. 496) This includes bug fixes to improve sorted set edge cases, accurate memory usage. For more information, see Redis 5.0.3 release notes. It also includes support for renaming commands. For more information, see ElastiCache for Redis Version 5.0.3 (Enhanced). February 28, 2019

ElastiCache Standard Reserved Instance offerings: Partial Upfront, All Upfront and No Upfront. (p. 496) 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. January 18, 2019

ElastiCache for Redis support for up to 250 nodes per Redis cluster The node or shard limit can be increased to a maximum of 250 per ElastiCache for Redis cluster. For more information, see Shards. November 19, 2018

ElastiCache for Redis support for autofailover and backup and restore on all T2 nodes (p. 496) ElastiCache for Redis introduces support for autofailover, creating snapshots, and backup and restore on all T2 nodes. For more information, see ElastiCache for Redis Backup and Restore and Snapshot. November 19, 2018
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support for ElastiCache for Redis 5.0.0</td>
<td>ElastiCache for Redis now supports Redis 5.0.0, including Redis streams. For more information, see <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/what-is-redis-compatibility.html">ElastiCache for Redis Version 5.0.0 (Enhanced)</a>. It has also added a new metric, <strong>StreamBasedCmds</strong>, which reports the sum of all of the commands that act upon one or more streams data type. For more information, see <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/what-is-redis.html">Metrics for Redis</a>.</td>
<td>November 9, 2018</td>
</tr>
<tr>
<td>ElastiCache for Redis support for M5 and R5 nodes</td>
<td>ElastiCache for Redis now supports M5 and R5 nodes, general-purpose and memory-optimized instance types based on the AWS Nitro System. For more information, see <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/elasti-cache-node-types.html">Supported Node Types</a>.</td>
<td>October 23, 2018</td>
</tr>
<tr>
<td>Support for dynamically changing the number of read replicas</td>
<td>ElastiCache for Redis has added support for adding and removing read replicas from any cluster with no cluster downtime. For more information about these and other changes in this release, see <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/changing-number-of-replicas.html">Changing the Number of Replicas</a> in the <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/">ElastiCache for Redis User Guide</a>. See also <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/API-2015-02-02/APIReference-Deletereplicacount.html">DecreaseReplicaCount</a> and <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/API-2015-02-02/APIReference-Increasereplicacount.html">IncreaseReplicaCount</a> in the <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/api-reference">ElastiCache API Reference</a>.</td>
<td>September 17, 2018</td>
</tr>
<tr>
<td>FedRAMP compliance certification</td>
<td>ElastiCache for Redis is now certified for FedRAMP compliance. For more information, see <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/fedramp.html">ElastiCache for Redis FedRAMP Compliance</a>.</td>
<td>August 30, 2018</td>
</tr>
<tr>
<td>Redis (cluster mode enabled) engine upgrades</td>
<td>Amazon ElastiCache for Redis has added support for upgrading Redis (cluster mode enabled) engine versions. For more information, see <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/">Upgrading Engine Versions</a>.</td>
<td>August 20, 2018</td>
</tr>
<tr>
<td>PCI DSS compliance certification</td>
<td>ElastiCache for Redis is now certified for PCI DSS compliance. For more information, see <a href="https://docs.aws.amazon.com/elasticache/latest/mdb/">ElastiCache for Redis PCI DSS Compliance</a>.</td>
<td>July 5, 2018</td>
</tr>
</tbody>
</table>
Support for ElastiCache for Redis 4.0.10

ElastiCache for Redis now supports Redis 4.0.10, including both encryption and online cluster resizing in a single version. For more information, see ElastiCache for Redis Version 4.0.10 (Enhanced).

June 14, 2018

User Guide restructure (p. 496)

The single ElastiCache User Guide is now restructured so that there are separate user guides for Redis (ElastiCache for Redis User Guide) and for Memcached (ElastiCache for Memcached User Guide). The documentation structure in the AWS CLI Command Reference: elasticache section and the Amazon ElastiCache API Reference remain unchanged.

April 20, 2018

Support for EngineCPUUtilization metric

ElastiCache for Redis added a new metric, EngineCPUUtilization, which reports the percentage of your CPU's capacity that is currently being used. For more information, see Metrics for Redis.

April 9, 2018

The following table describes the important changes to the ElastiCache for Redis User Guide before March 2018.

<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
</table>
| 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. 56)  
  - Supported Node Types (p. 65) | February 12, 2018 |
| Support for EU (Paris). | ElastiCache added support for the EU (Paris) Region. For more information, see the following:  
  - Supported Regions & Endpoints (p. 56)  
  - Supported Node Types (p. 65) | December 18, 2017 |
| Support for China (Ningxia) Region | Amazon ElastiCache added support for China (Ningxia) Region. For more information, see the following:  
  - Supported Regions & Endpoints (p. 56)  
  - Supported Node Types (p. 65) | December 11, 2017 |
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
</table>
| 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. 401)  
  - Set Up Your Permissions (New ElastiCache Users Only) (p. 25) | December 7, 2017 |
| Support for R4 node types | This release of ElastiCache added support for 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. 65)  
  - Redis Node-Type Specific Parameters (p. 336) | November 20, 2017 |
| ElastiCache for Redis 3.2.10 and support for online resharding | Amazon ElastiCache for Redis adds support for ElastiCache for Redis 3.2.10. ElastiCache for Redis also introduces online cluster resizing to add or remove shards from the cluster while it continues to serve incoming I/O requests. For more information, see the following:  
  - Best Practices: Online Cluster Resizing (p. 472)  
  - Online Resharding and Shard Rebalancing for Redis (cluster mode enabled) (p. 281) | November 9, 2017 |
| HIPAA eligibility | ElastiCache for Redis version 3.2.6 is now certified for HIPAA eligibility when encryption is enabled on your cluster. For more information, see the following:  
  - HIPAA Eligibility (p. 452)  
  - Data Security in Amazon ElastiCache (p. 340) | November 2, 2017 |
| ElastiCache for Redis 3.2.6 and support for encryption | ElastiCache adds support for ElastiCache for Redis 3.2.6, which includes two encryption features:  
  - In-transit encryption encrypts your data whenever it is in transit, such as between nodes in a cluster or between a cluster and your application.  
  - At-rest encryption encrypts your on-disk data during sync and backup operations.  
  For more information, see the following:  
  - Data Security in Amazon ElastiCache (p. 340)  
  - Supported ElastiCache for Redis Versions (p. 45) | October 25, 2017 |
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection patterns topic</td>
<td>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. 363) in the ElastiCache User Guide.</td>
<td>April 24, 2017</td>
</tr>
</tbody>
</table>
| Support for testing Automatic Failover | ElastiCache adds support for testing Automatic Failover on Redis clusters that support replication. For more information, see the following:  
  - TestFailover in the ElastiCache API Reference.  
  - test-failover in the AWS CLI Reference. | April 4, 2017       |
| Enhanced Redis restore         | ElastiCache adds enhanced Redis backup and restore with cluster resizing. This feature supports restoring a backup to a cluster with a different number of shards than the cluster used to create the backup. (For the API and CLI, this feature can restore a different number of node groups rather than a different number of shards.) This update also supports different Redis slot configurations. For more information, see Restoring From a Backup with Optional Cluster Resizing (p. 239). | March 15, 2017     |
| New Redis memory management parameter | ElastiCache adds a new Redis parameter, reserved-memory-percent, which makes managing your reserved memory easier. This parameter is available on all versions of ElastiCache for Redis. For more information, see the following:  
  - Managing Reserved Memory (p. 466)  
  - New Parameters for Redis 3.2.4 (p. 323) | March 15, 2017     |
| 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. 56)  
  - Supported Node Types (p. 65) | 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. 56)  
  - Supported Node Types (p. 65) | December 8, 2016    |
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
</table>
| 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. 56)  
- Supported Node Types (p. 65) | November 1, 2016 |
| 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. 56)  
- Supported Node Types (p. 65) | October 17, 2016 |
| Support for Redis Cluster | ElastiCache adds support for Redis Cluster (enhanced). Customers using Redis Cluster, can partition their data across up to 15 shards (node groups). Each shard supports replication with up to 5 read replicas per shard. Redis Cluster automatic failover times are about one fourth as long as those of earlier versions.  
This release includes a redesigned management console that uses terminology in keeping with industry usage.  
For more information, see the following:  
- Comparing Memcached and Redis  
- ElastiCache for Redis Components and Features (p. 10) — note the sections on Nodes, Shards, Clusters, and Replication.  
| 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. 65)  
- Redis Node-Type Specific Parameters (p. 336) | 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. 65)  
- Redis Node-Type Specific Parameters (p. 336) | June 27, 2016 |
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
</table>
| Snapshot export | ElastiCache adds the ability to export a Redis snapshot so you can access it from outside ElastiCache. For more information, see the following:  
  - Exporting a Backup (p. 231) in the Amazon ElastiCache User Guide  
  - CopySnapshot in the Amazon ElastiCache API Reference | May 26, 2016 |
| Node type scale up | ElastiCache adds the ability to scale up your Redis node type. For more information, see Scaling ElastiCache for Redis Clusters (p. 250). | March 24, 2016 |
| Easy engine upgrade | ElastiCache adds the ability to easily upgrade your Redis cache engine. For more information, see Upgrading Engine Versions (p. 53). | March 22, 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. 65). | 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. 34). | February 12, 2016 |
| Support for Redis 2.8.24 | ElastiCache adds support for Redis version 2.8.24 with improvements added since Redis 2.8.23. Improvements include bug fixes and support for logging bad memory access addresses. For more information, see the following:  
  - ElastiCache for Redis Version 2.8.24 (Enhanced) (p. 51)  
  - Redis 2.8 Release Notes | January 20, 2016 |
<p>| 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 |
| Amazon ElastiCache console change. | Because the newer Redis versions provide a better and more stable user experience, Redis versions 2.6.13, 2.8.6, and 2.8.19 are no longer listed in the ElastiCache Management Console. For other options and more information, see Supported ElastiCache for Redis Versions (p. 45). | December 15, 2015 |
| Support for Redis 2.8.23 | ElastiCache adds support for Redis version 2.8.23 with improvements added since Redis 2.8.22. Improvements include bug fixes and support for the new parameter close-on-slave-write which, if enabled, disconnects clients who attempt to write to a read-only replica. For more information, see ElastiCache for Redis Version 2.8.23 (Enhanced) (p. 51). | November 13, 2015 |</p>
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Implementation of a forkless save process that enables a successful save when low available memory could cause a forked save to fail.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Additional CloudWatch metrics — <code>SaveInProgress</code> and <code>ReplicationBytes</code>.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• To enable partial synchronizations, the Redis parameter <code>repl-backlog-size</code> now applies to all clusters.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>For a complete list of changes and more information, see ElastiCache for Redis Version 2.8.22 (Enhanced) (p. 51).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>This documentation release includes a reorganization of the documentation and removal of the ElastiCache command line interface (CLI) documentation. For command line use, refer to the AWS Command Line for ElastiCache.</td>
<td></td>
</tr>
<tr>
<td>Support for Redis 2.8.21</td>
<td>ElastiCache adds support for Redis version 2.8.21 and Redis improvements since version 2.8.19. This Redis release includes several bug fixes. For more information, see Redis 2.8 release notes.</td>
<td>July 29, 2015</td>
</tr>
<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. 123).</td>
<td>July 9, 2015</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 Redis v. 2.8.19. | ElastiCache adds support for Redis version 2.8.19 and Redis improvements since version 2.8.6. This support includes support for:  
- The HyperLogLog data structure, with the Redis commands PFADD, PFCOUNT, and PFMERGE.  
- Lexicographic range queries with the new commands ZRANGEBYLEX, ZLEXCOUNT, and ZREMRANGEBYLEX.  
- Introduced a number of bug fixes, namely preventing a primary node from sending stale data to replica nodes by failing the master SYNC when a background save (bgsave) child process terminates unexpectedly.  
For more information on HyperLogLog, see Redis new data structure: the HyperLogLog.  
For more information on PFADD, PFCOUNT, and PFMERGE, see the Redis Documentation and click HyperLogLog. | March 11, 2015 |
<p>| Support for cost allocation tags | ElastiCache adds support for cost allocation tags. For more information, see Monitoring Costs with Cost Allocation Tags (p. 432). | February 9, 2015 |
| Support for Europe (Frankfurt) Region | ElastiCache adds support for the Europe (Frankfurt) (eu-central-1) Region. | January 19, 2015 |
| Multi-AZ with auto failover support for Redis replication groups | ElastiCache adds support for Multi-AZ with automatic failover from the primary node to a read replica in a Redis replication group. ElastiCache monitors the health of the replication group. If the primary fails, ElastiCache automatically promotes a replica to primary, then replaces the replica. For more information, see Minimizing Downtime: Multi-AZ with Automatic Failover (p. 148). | October 24, 2014 |
| AWS CloudTrail logging of API calls supported | 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. 486). | September 15, 2014 |
| New instance sizes supported | ElastiCache adds support for additional General Purpose (T2) instances. For more information, see Configuring Engine Parameters Using Parameter Groups (p. 298). | September 11, 2014 |</p>
<table>
<thead>
<tr>
<th>Change</th>
<th>Description</th>
<th>Date Changed</th>
</tr>
</thead>
<tbody>
<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. 298).</td>
<td>July 1, 2014</td>
</tr>
<tr>
<td>Backup and restore for Redis clusters</td>
<td>In this release, ElastiCache allows customers to create snapshots of their Redis clusters, and create new clusters using these snapshots. A backup is a copy of the cluster at a specific moment in time, and consists of cluster metadata and all of the data in the Redis cache. Backups are stored in Amazon S3, and customers can restore the data from a snapshot into a new cluster at any time. For more information, see Backup and Restore for ElastiCache for Redis (p. 213).</td>
<td>April 24, 2014</td>
</tr>
<tr>
<td>Redis 2.8.6</td>
<td>ElastiCache supports Redis 2.8.6, in addition to Redis 2.6.13. With Redis 2.8.6, customers can improve the resiliency and fault tolerance of read replicas, with support for partial resynchronization, and a user-defined minimum number of read replicas that must be available at all times. Redis 2.8.6 also offers full support for publish-and-subscribe, where clients can be notified of events that occur on the server.</td>
<td>March 13, 2014</td>
</tr>
<tr>
<td>Redis cache engine</td>
<td>ElastiCache offers Redis cache engine software, in addition to Memcached. Customers who currently use Redis can “seed” a new ElastiCache Redis cache cluster with their existing data from a Redis snapshot file, easing migration to a managed ElastiCache environment. To support Redis replication capabilities, the ElastiCache API now supports replication groups. Customers can create a replication group with a primary Redis cache node, and add one or more read replica nodes that automatically stay synchronized with cache data in the primary node. Read-intensive applications can be offloaded to a read replica, reducing the load on the primary node. Read replicas can also guard against data loss in the event of a primary cache node failure.</td>
<td>September 3, 2013</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. 358).</td>
<td>January 8, 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. 358).</td>
<td>December 20, 2012</td>
</tr>
<tr>
<td>Change</td>
<td>Description</td>
<td>Date Changed</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------</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 <em>Amazon ElastiCache User Guide</em>.</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.