

Amazon ElastiCache for Redis

User Guide

▶ What is ElastiCache for Redis?

▶ Getting started with ElastiCache for Redis

▶ Designing your own ElastiCache cluster

▼ Working with ElastiCache

▶ Snapshot and restore

▶ Engine versions and upgrading

▶ Best practices and caching strategies

▶ Managing your self-designed cluster

▼ Scaling ElastiCache for Redis

▼ Scaling ElastiCache for Redis self-designed clusters

▶ Scaling clusters for Redis (Cluster Mode Disabled)

▼ **Scaling clusters in Redis (Cluster Mode Enabled)**

Offline resharding and shard rebalancing for Redis (cluster mode enabled)

Online resharding and shard rebalancing for Redis (cluster mode enabled)

▶ Online vertical scaling by modifying node type

▶ Getting started with JSON in ElastiCache for Redis

▶ Tagging your ElastiCache resources

▶ Amazon ElastiCache Well-Architected Lens

Troubleshooting

Additional troubleshooting steps

▶ Security

▶ Logging and monitoring

Quotas

▶ Reference

ElastiCache for Redis

Documentation history

AWS Glossary

# Scaling clusters in Redis (Cluster Mode Enabled)

[PDF](#) | [RSS](#)

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 [Select cache node size](#).

## Contents

- [Offline resharding and shard rebalancing for Redis \(cluster mode enabled\)](#)
- [Online resharding and shard rebalancing for Redis \(cluster mode enabled\)](#)
  - [Adding shards with online resharding](#)
  - [Removing shards with online resharding](#)
    - [Removing shards \(Console\)](#)
    - [Removing shards \(AWS CLI\)](#)
    - [Removing shards \(ElastiCache API\)](#)
  - [Online shard rebalancing](#)
    - [Online Shard Rebalancing \(Console\)](#)
    - [Online shard rebalancing \(AWS CLI\)](#)
    - [Online shard rebalancing \(ElastiCache API\)](#)
- [Online vertical scaling by modifying node type](#)
  - [Online scaling up](#)
    - [Scaling up Redis cache clusters \(Console\)](#)
    - [Scaling up Redis cache clusters \(AWS CLI\)](#)
    - [Scaling up Redis cache clusters \(ElastiCache API\)](#)
  - [Online scaling down](#)
    - [Scaling down Redis cache clusters \(Console\)](#)
    - [Scaling down Redis cache clusters \(AWS CLI\)](#)
    - [Scaling down Redis cache clusters \(ElastiCache API\)](#)

## View related pages [Preview](#)

< 1 2 3 4 5 6 >

Memorydb › devguide

Scaling MemoryDB clusters

January 19, 2024

Memorydb › devguide

Online vertical scaling by modifying node type

January 19, 2024

## Discover highly rated pages [Preview](#)

< 1 2 3 4 5 >

AmazonElastiCache › red-ug

What is Amazon ElastiCache for Redis?

February 7, 2024

AmazonElastiCache › mem-ug

What is Amazon ElastiCache for Memcached?

February 7, 2024

Did this page help you?

[Provide feedback](#)

**Next topic:** [Offline resharding and shard rebalancing for Redis \(cluster mode enabled\)](#)

**Previous topic:** [Decreasing read capacity](#)

### Need help?

- [Try AWS re:Post](#)
- [Connect with an AWS IQ expert](#)