



ElastiCache for RDS Deployment option

1

Introduction

1

After selecting the cluster type (Redis or Memcached) in Amazon ElastiCache for RDS, the next option you'll see is the Deployment Mode.

2

This step helps you decide how the cache will be set up to manage data and ensure availability

3

There are two options

1

Serverless Cache

2

Design Your Own Cache

2

Serverless Cache

1

Introduction

1

This is a fully managed option where AWS automatically handles everything for you

2

This option is supported by both Redis and Memcached

2

Key Features

1

Automatic Scaling Resources scale up or down based on your application's needs

2

No Configuration Needed You don't need to manually set up nodes, shards, or replicas

3

Pay-as-You-Go You only pay for the cache resources you use

3

Best For

1

Applications with unpredictable or fluctuating traffic

2

Quick and Hassle-Free Deployments:

3

Scalable Applications Without Management Overhead:

3

Design Your Own Cache

1

Introduction

1

The Design Your Own Cache option provides full control over your cache setup.

2

This option allows you to customize the number of nodes, shards, replicas, and other advanced settings.

3

Additionally, you can select between

1

Cluster Mode Enabled (Supported only by Redis)

2

Cluster Mode Disabled (Supported by both Redis and Memcached)

2

Key Features

1

Advanced Control and Customization

1

Gain full flexibility to design your cache infrastructure to suit your specific workload requirements.

2

You can tailor performance, cost, and scalability by choosing the node types, configuring replicas, and defining the overall architecture

2

Cost Optimization

Design an architecture that balances performance and cost, such as minimizing replicas for smaller workloads or leveraging advanced node types for higher efficiency in larger deployments

3

Controlled Scalability

1

While Serverless Cache automatically scales resources without user input, Design Your Own Cache allows you to define and control the scalability of your infrastructure.

2

You can decide the number of nodes, shards, and replicas to match your application's specific performance and capacity requirements

3

Best For

1

Applications with Predictable Workloads

2

Use Cases Requiring Custom Infrastructure

3

Cost-Sensitive Deployments