

Auto Scaling Tutorial Product Documentation



Copyright Notice

©2013-2019 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Tutorial

Creating Web Services

Creating High-performance Computing Cluster

Creating Servers for Sending Requests

Configuring For High Availability Services

Tutorial

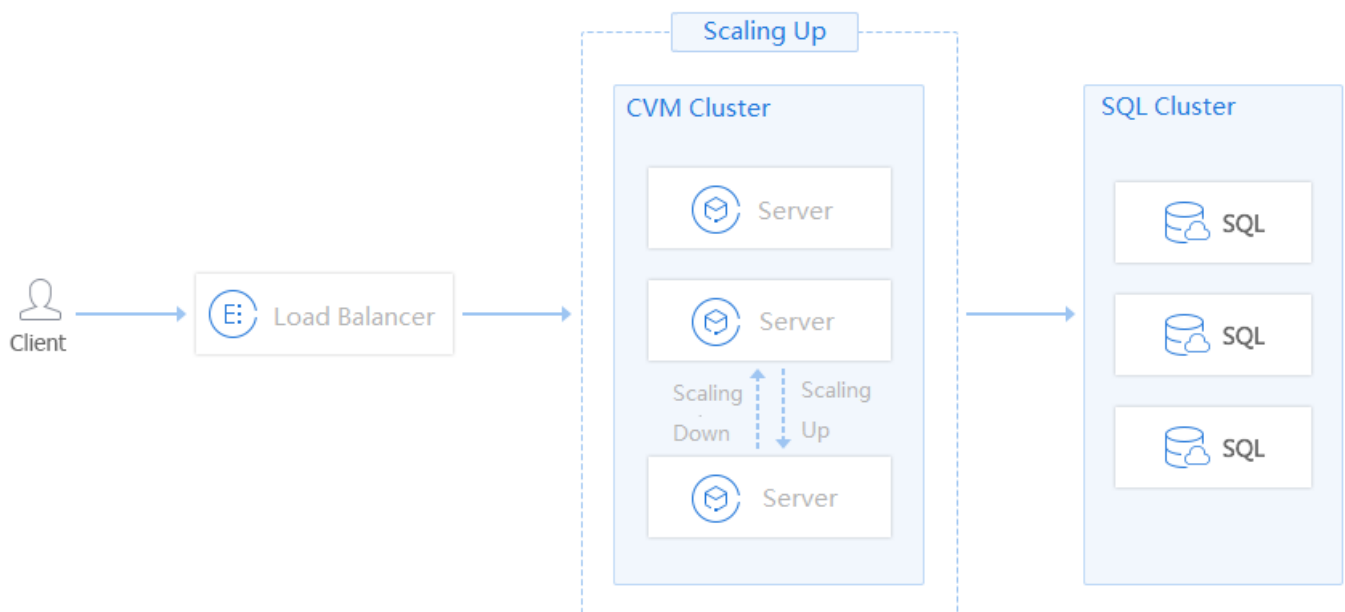
Creating Web Services

Last updated : 2019-12-30 16:48:29

Use Cases

Auto Scaling (AS) is applicable to servers used for common Web services. It helps you to manage servers easily and increase the business robustness.

For e-commerce websites, video websites, and online education websites, requests from clients are directs to the application server through load balancers. In case of rapid changes in the access volume, Auto Scaling can flexibly scale in and out application CVMs based on the volume of requests.



Usage

Add the following clusters to the scaling group to provide further protection for the cluster:

- Frontend server cluster (access layer)
- Application server cluster (logic layer)
- Cache server cluster (data layer)

You can also configure scheduled scaling policy for expected business peaks (such as online promotions).

Add CVMs in the cluster to the scaling group, and enable **Scale-in Protection** for certain CVMs to ensure the normal operation of the cluster. You can also configure the alarm scaling policy to prevent burst traffic or CC attacks.

Benefits of AS

1. AS can help deal with unexpected request traffic and avoid single point of failures.
2. You only need to estimate the resident resources, AS can dynamically adjust elastic resources to reduce IT costs.
3. Scale out quickly in case of CC attacks to avoid request packet loss.

Applicable Industries

This solution is applicable to all websites, especially to those with large load fluctuations:

- E-commerce websites
- Online education websites
- Video websites
- Live streaming websites

FAQs

Is this solution applicable to common Web services, such as internal systems or websites with steady traffic?

Yes. Common websites will also encounter unexpected situations, such as CC attacks, or piled access due to unexpected incidents.

This solution does no incur any additional cost. You only need to enable **Scale-in Protection** for the necessary servers, and don't need to change the structure and operation of websites.

It's recommended to take this solution as the best practice for your websites.

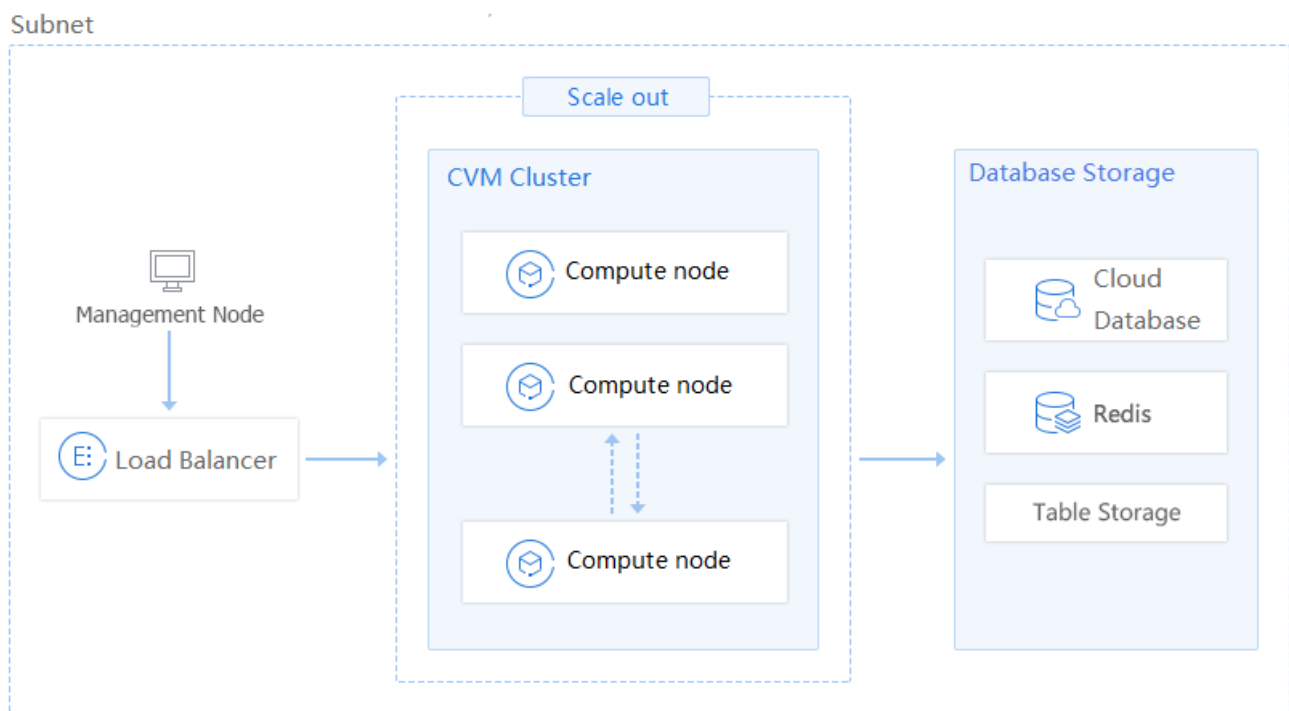
Creating High-performance Computing Cluster

Last updated : 2019-12-17 17:04:18

Use Case

With cloud computing, high performance computing (HPC) can use applications with higher bandwidth and higher computing capacity to address complex scientific, engineering and business issues.

However, the problems solved by HPC are usually based on projects, and have huge demand for the high scalability of the cloud platform. The compute node can be configured in a launch configuration (template) for the scaling group. By increasing the desired capacity, multiple compute nodes can be generated with one click for any calculation tasks. After saving the calculation results, you can delete the compute nodes for the task by modifying the desired capacity.



Usage

Create a launch configuration for the nodes in the cluster, and place the computing cluster into the scaling group.

There are two ways to use the original data for high performance computing:

- Save the data to a snapshot, so that the data disks of the scaled-out CVMs are created based on the snapshot.
- Save the data to a data server, so that all the compute nodes in the cluster can be read in the data server.

Benefits of AS

- AS can greatly reduce the workload of manual preparation of environments.
- There is no need to reserve long-term resources for temporary tasks.

Applicable industries

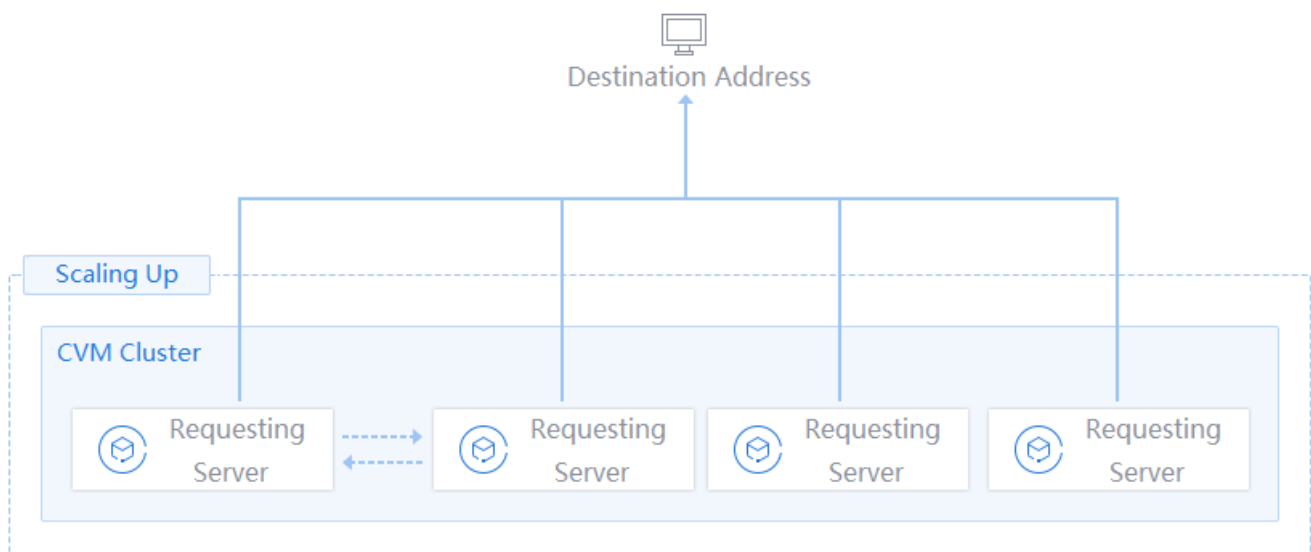
- Weather forecasting
- Gene sequencing
- Animation rendering, film and TV rendering
- Other industries that require high performance computing

Creating Servers for Sending Requests

Last updated : 2019-12-16 16:00:32

The active outbound CVMs require timeliness. In this case, the Auto Scaling service can facilitate the rapid creation, deployment, and scale-in of the request CVMs.

If you have a clear idea about the time need to scale in an out, you can set a scheduled scaling policy to automatically add or remove CVM instances at the specified time, eliminating the need for manual operation and reducing your deployment and instance costs.



Configuring For High Availability Services

Last updated : 2020-02-11 20:04:32

Use case

Building master/slave or active-active HA clusters in the traditional way is quite cumbersome. You can use health check of Auto Scaling to implement high availability.

The system will automatically monitor the health status of the active nodes. When the active node does not respond to a ping, Auto Scaling will automatically replicate a healthy instance to replace any abnormal ones, ensuring healthy and smooth business operation and providing comprehensive protection for your business.

Directions

Step 1: Create images of stateless CVMs in a cluster.

Step 2: Create a scaling group, and configure the maximum and the minimum capacities as desired. Then, select **Add a CVM** on the list of CVMs in the scaling group to manually add the existing CVM to the cluster. Note: When a manually added CVM is replaced, it's only removed from the scaling group but not terminated.

Step 3: Create a notification policy and select to receive notifications about scaling activities that replace unhealthy instances.

New Notice ✕

Current scaling group

- Expansion succeeded
- Expansion failed
- Capacity reduction succeeded
- Capacity reduction failed
- Unhealthy instance replaced successfully
- Replacing unhealthy instance failed

Send notice to

+ Create a new user group that receives notifications

yunyxiao-group

test

[How to define a user group that receives notifications](#)

The configuration is complete.

Benefits of AS

AS can make the cluster secure.

Applicable industries

It is strongly recommended that you add stateless CVMs (if any) to the scaling group, and make this an IT deployment routine.