Auto Scaling

Tutorial

Product Introduction

Tencent
Cloud

Copyright Notice

Trademark Notice

Tencent Cloud

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

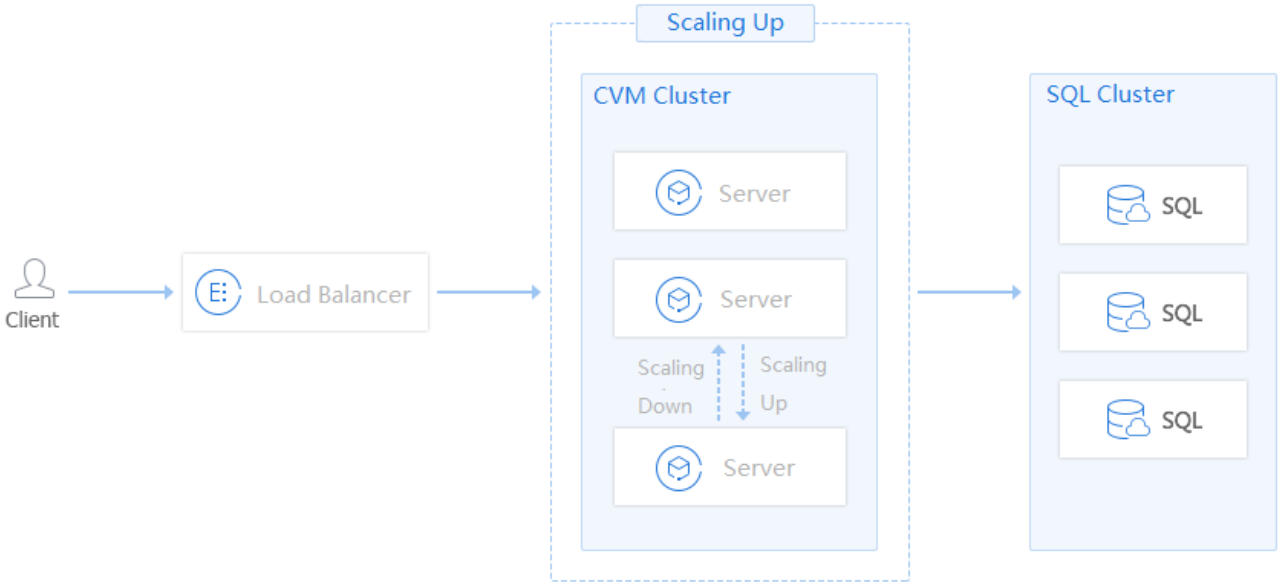
Tutorial

Creating Web Services

Scenario Description

It is recommended to use AS (Auto Scaling) for common Web service server to get free access to convenient management and robust business.

For e-commerce websites, video websites and online education applications, client requests are sent to application server through cloud load balancer. In case of rapid changes in visits, AS service can flexibly scale up and down the number of application servers based on the number of requests.



How to Use

Add the following clusters to a 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 set up timed scaling policies for expected business peaks (such as online promotions).

> TIPS: Move the CVMs in the cluster into the scaling group, and set the "scale-down exemption" for certain CVMs to ensure normal use of the cluster. At the same time, set the alarm scaling policy to cater with burst traffic or CC attacks.

Benefits of AS

1. AS can provide further protection for your business by dealing with unexpected request traffic and avoiding SPOF;

2. By only estimating resident resources, instead of estimating resources based on the peak value, AS can dynamically adjust the elastic resources to save IT cost;

3. Rapid scale-up can be enabled in case of CC attacks to avoid request packet loss.

Applicability

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

- E-commerce website
- Online education website
- Video website
- LVB website

FAQ

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

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

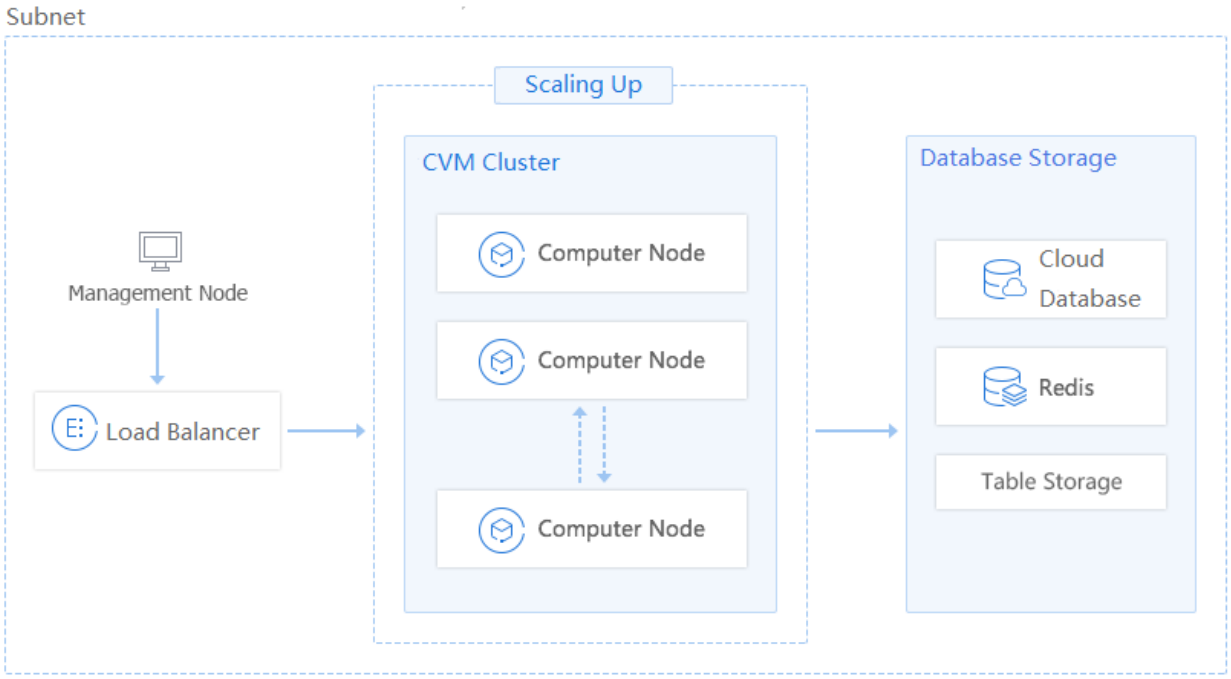
The solution will not cause any additional cost. By simply setting "scale-down exemption" for the planned CVM, the design and operation of websites will not be affected. When any accidents occur, AS can bring huge benefits, avoiding service suspension.

Therefore, we highly recommend this solution for you.

Creating High-performance Computing Cluster

Scenario Description

Cloud computing enables high performance computing (HPC) to use applications with higher bandwidth and higher computing capacity to address complex scientific, engineering and business issues.
But the problems solved by HPC are usually based on projects, with huge demands for the high scalability of the cloud platform. Compute node can be set into a scaling configuration (template) for the scaling group. By increasing the desired instance number, multiple compute nodes will 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 instance number.



Tips on Usage

Create a scaling 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 into snapshot, so that the CVM's expanded data disk is created based on the snapshot;
- Save the data into data server, so that all the compute nodes in the CVM can be read in the data server.

Benefits of AS

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

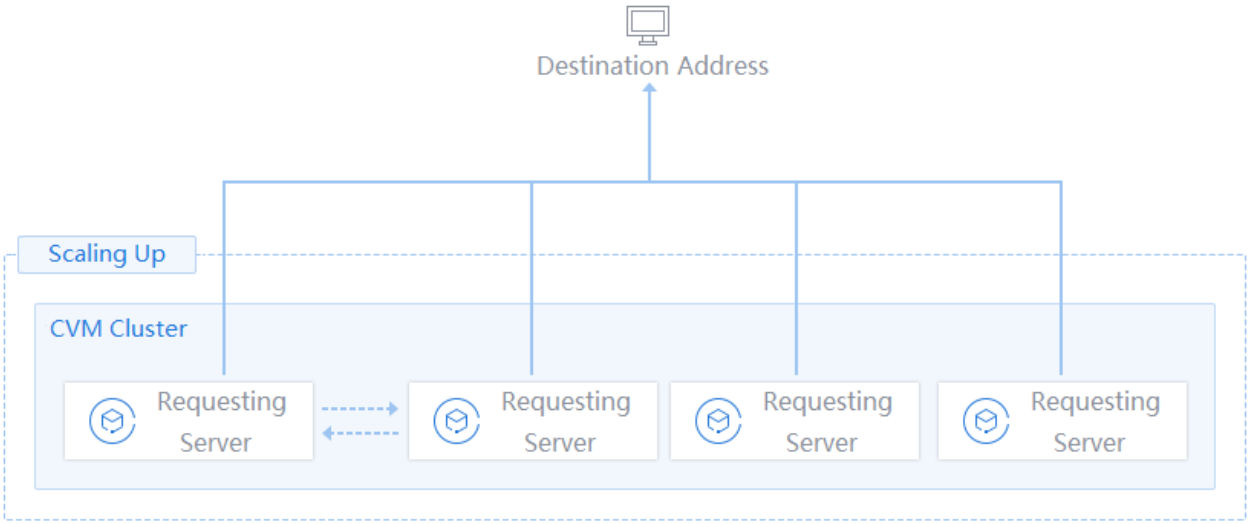
Applicability

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

Creating Servers for Sending Requests

The server accessing externally is time-efficient. In this case, Auto Scaling service can facilitate fast creation, deployment and scale-down of the server sending requests.

If you know when capacity scaling is needed, you can configure Auto Scaling scheduled policy in advance. By the configured time, the system will automatically increase or decrease the number of CVM instances without waiting, saving your deployment and instance costs.



Configuring For High Availability Services

Scenario Description

It is relatively cumbersome to build traditional master/slave or active-active HA clusters. You can use health check of Auto Scaling to achieve high availability.

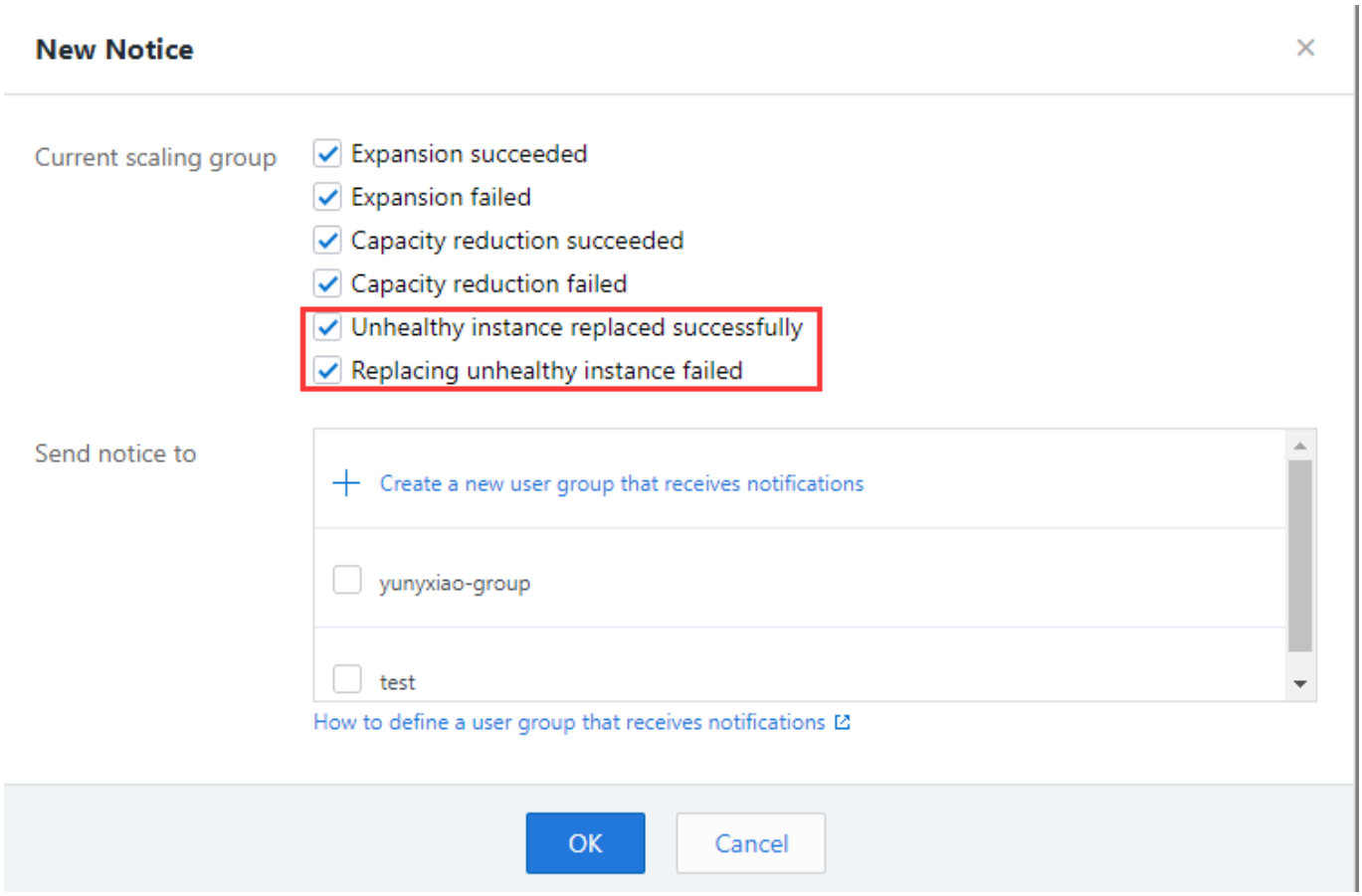
The system will automatically monitor the health status of the active nodes. When the active node does not respond to a ping, the Auto Scaling will automatically replicate a healthy instance to replace any abnormal ones, to ensure healthy and smooth business operation and provide all-round protection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

Tips on Usage

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

Step 2: Create a scaling group, and set the maximum and the minimum scaling group sizes. After that, select Add CVM from the list of CVMs in the scaling group to manually add the existing CVM in the cluster. Note: When a CVM that is manually added to a scaling group is replaced, such CVM is not destroyed, but only removed from the scaling group.

Step 3: Create a notification and select to accept the notification on the scaling activities that replace unhealthy instances



Benefits of AS

AS can help secure the cluster.

Applicability

It is strongly recommended that you add the stateless CVM (if any) to the scaling group, as a routine IT deployment.