

Auto Scaling

FAQs

Product Documentation



Copyright Notice

©2013-2024 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

 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.

FAQs

Last updated : 2024-01-08 17:53:30

What is the cooldown period?

The cooldown period refers to the locking duration after one scaling activity (the addition or removal of CVM instances) is executed in a scaling group. During which, the scaling group will not perform any scaling activities. You can optionally set the cooldown period to a value less than 999,999 seconds.

Do manually added CVM instances have a cooldown period?

No.

Is there a charge for Auto Scaling?

No. Please rest assured that the Auto Scaling feature is completely free of charge.

You only need to pay for other services associated with AS, such as CVM instances automatically created by AS or those manually added. For billing details, see [Billing Plans](#).

How do I increase the max capacity of a scaling group?

A scaling group can have up to 2,000 CVM instances. For more information about the current quota of pay-as-you-go CVM instances under an account in each AZ, see [Purchase Limits](#). If you want to have more than 2,000 CVM instances in a scaling group, or if you want to use more pay-as-you-go CVM instances than the fixed quota, [submit a ticket](#) for application.

What kinds of servers are suitable for Auto Scaling?

Applications that are deployed in the CVM instances in the scaling group should be stateless and replaceable. Because instances in the scaling group are likely to be repossessed by Tencent Cloud during scale-in, CVM instances for Auto Scaling cannot save the status information (such as sessions) and data (such as databases and logs) of the applications. If you need to save the status information in your application, you can consider saving it in a standalone CVM outside the scaling group.

What kinds of CVMs can be manually added to a scaling group?

A CVM that is manually added to the scaling group must meet the following requirements:

In the same region as the scaling group.

In the same network environment (VPC or classic network) as the scaling group.

In the running status.

What kinds of CLB instances can be associated with the scaling group?

A CLB instance (in the case of a cross-region CLB instance, its real server) that is associated with a scaling group must be in the same network environment (VPC or the classic network of the same region) as the scaling group.

Can Auto Scaling automatically upgrade or degrade the configurations of CVMs?

No. Auto Scaling is a management service that elastically adjusts computing resources according to your business needs and policies. It can automatically add CVM instances when the business grows and remove CVM instances when the business declines. Currently, Auto Scaling does not support "scaling up", that is, Auto Scaling cannot automatically change the CPU, memory and bandwidth of CVM instances.

Does Auto Scaling have to work with CLB and Cloud Monitor?

No. Auto Scaling can independently scale CVM instances out and in. It can work with CLB or work alone.

What should I do to add CVMs during a specific period of time?

You can configure two scheduled actions: one defines the desired capacity of scale-out, and the other defines the desired capacity of scale-in.

What are the conventions for naming an automatically added CVM instance?

The scaling group will name an automatically added CVM instance in the format of `as-scaling group name+auto-incrementing numbers`.

For example, if your scaling group is named `test` and two CVM instances are automatically added, the instances will be named `as-test1` and `as-test2` respectively.

What are the specific rules for removing CVMs from the scaling group?

Tencent Cloud Auto Scaling provides two removal policies:

Remove the oldest instances: removes the earliest auto-added instances from the scaling group. Instances added automatically are removed first, and then the oldest instances that are manually added.

Remove the latest instances: removes the latest auto-added CVMs from the scaling group. Instances added automatically are removed first, and then the latest instances that are manually added.

How does the alarm policy collect the Cloud Monitor data?

Take the maximum value as an example. It collects one sample per minute from each CVM in the period. If the sample value meets the configured rules for multiple consecutive periods (custom), the alarm-based scaling action will be triggered.

For example, there are five CVMs in a scaling group, and the defined alarm scaling policy is "CPU utilization exceeds 50% for 3 consecutive periods". The system takes one value per minute from each CVM, that is, it takes 25 CPU utilization values within one period (5 minutes). If there are values exceeding 50% for 3 consecutive periods, the scaling rule is satisfied and scaling actions are triggered.

What is the desired capacity?

The desired capacity is the reasonable number of instances in the current scaling group. Its value should be between the min capacity and the max capacity. You can manually adjust the desired capacity, and you can also use scheduled action and alarm-based scaling task to trigger adjustment. The scaling group automatically adjusts the actual number of instances to match the desired capacity.

Configure on scaling group creation: if you set an initial capacity when creating the scaling group, then the desired capacity will be the initial capacity.

Adjusted by alarm-based scaling task: when alarm-based scaling is triggered, the scaling group will adjust the current capacity to the desired capacity. For example, if the triggered action is to add two CVMs, the backend will add two CVMs to the desired capacity to achieve this, and then add two instances to the current capacity to make them equal.

Adjusted manually or by scheduled actions: if the desired capacity was modified manually or through a scheduled action, the system will trigger a scaling action to make the current capacity equal to the desired capacity.

Automatically adjusted: the desired capacity should be between the max capacity and the min capacity. If either max or min capacity changes, the desired capacity may change accordingly. For example, assume that the desired capacity is 3, the min capacity is 2 and the max capacity is 5. If the min capacity changes to 4, then the desired capacity will be adjusted to 4 to be no less than the min capacity.

What should be noted when a data disk snapshot is specified in the launch configuration?

If the launch configuration specifies a data disk snapshot, you must ensure that the data disk can be automatically attached to ensure proper auto scale-out. Please configure the source instance of the snapshot as instructed below: [Attaching Cloud Disks](#).

What will happen when the scaling group is disabled?

After a scaling group is disabled, the automatically triggered activities will not proceed.

Automatically triggered activities include:

Alarm-based scaling

Scheduled actions

Health checks

Matching the current capacity due to manual operations with the desired capacity

Scaling group restrictions include:

If manual addition causes the current capacity to be more than the max capacity, the addition will not be permitted.

Modifying the min or max capacity of the scaling group does not trigger a scaling action, but the modification takes effect.

Manual removal of instances is not limited by the min capacity.

What stages are involved in the lifecycle of the CVMs that are automatically added to a scaling group?

Creating: the CVM is being created

InService: the CVM is in service

Removing: the CVM is being removed

Attaching: the CVM is being bound to the scaling group

Detaching: the CVM is being unbound from the scaling group

AttachLb: the CVM is being bound to a CLB

DetachLb: the CVM is being unbound from a CLB

Preheating: a CVM is preheating

What are the rules for CVM removal?

Removal of a manually added CVM: the CVM is only removed from the scaling group but not terminated. This removal will disassociate the CVM from the CLB instance that has been automatically bound in the scaling group, rather than the CLB instance manually bound.

Removal of an automatically created CVM: the CVM will be terminated and unbound from the CLB.