

Auto Scaling

FAQs

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

FAQs

Elastic Scaling Issues

FAQs

Elastic Scaling Issues

Last updated : 2020-03-04 17:38:01

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. The scaling group does not perform scaling activities during this duration. The value range of the cooldown period is 0 - 999999 (seconds).

Do manually added CVMs have a cooldown period?

No. Manually added CVMs do not need a cooldown period.

Is there a charge for Auto Scaling?

No. The Auto Scaling feature is completely free of charge. Please feel free to use it. CVM instances that are automatically created through auto scaling are billed in normal pay-as-you-go mode. Manually added CVM instances keep their original billing policy, and this is not affected by scaling activities.

How do I increase the maximum number of CVMs in the scaling group?

At most 2000 CVMs are supported in one scaling group. For more information about the current quota of pay-as-you-go CVMs that Tencent Cloud users can have in each availability zone, see [Purchase Limits for Pay-as-You-Go CVM Instances](#). If you want to have more than 2000 CVMs in one scaling group, or if you want to use more pay-as-you-go CVMs than the fixed quota, [submit a ticket](#).

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 related data (such as databases, logs, etc.) of the applications. If you need to save the status information in your application, you can consider saving the status information in a standalone CVM outside the scaling group.

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

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

- Locate in the same region as the scaling group.
- Be in the same network environment (VPC or basic network) as the scaling group.

- Be in a normal running status.

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

A CLB instance that is associated with a scaling group must be in the same network environment (VPC or the basic 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 can scale in CVM instances when business decreases. Currently, Auto Scaling does not support "vertical scaling", that is, Auto Scaling cannot automatically change the CPU, memory and bandwidth of CVMs.

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

No. Auto Scaling can separately scale CVM instances out and in, and it can work with CLB for deployment or can work on its own.

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

You can configure two scheduled actions: one defines the desired number of instances of scale-out. The other defines the desired number of instances of scale-in.

What are the specific rules for the scaling group removal policies?

Tencent Cloud Auto Scaling provides two removal policies:

- Remove the oldest: Remove the oldest instances in the group. Instances added automatically are removed first, and then the manually added ones.
- Remove the newest: Remove the newest instances in the group. Instances added automatically are removed first, and then the manually added ones.

How does the alarm policy collect the cloud monitoring information?

Take the maximum value as an example. It collects one sample per minute for each of CVM in the period. If the sample value meets the configured rules for multiple periods consecutively (the number of periods can be customized), the alarm scaling activity 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 exceed 50% for 3 consecutive periods, the scaling rule is satisfied and scaling actions are triggered..

What is desired capacity?

The desired capacity is the reasonable number of instances in the current scaling group, and the size is between the minimum capacity and the maximum capacity. You can manually adjust the desired capacity, and you can also use scheduled action and alarm scaling tasks 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 scaling task: when alarm scaling is triggered, the scaling group will adjust the current number of instances to the desired capacity. For example, if the triggered action is to add two CVMs, the backend will add two CVMs to the current number to achieve this. If the system finds that the current number of instances of the scaling group is not equal to the desired capacity, it will add two CVMs to make them equal.
- Adjusted manually or by scheduled actions: if the desired capacity was modified manually or through a scheduled action, and the system finds that the current number of instances is not equal to the desired capacity, the system will trigger a scaling action to make the number equal to the desired capacity.
- Automatically adjusted: the desired capacity should be between the maximum and minimum capacity. If the maximum or minimum capacity changes, this may result in a change to the desired capacity. For example, assume that the desired capacity is 3, the minimum capacity is 2 and the maximum capacity is 5. If the minimum capacity is adjusted to 4, then the desired capacity will be adjusted to 4 to meet the minimum 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 mounted. Before configuring auto scaling, you must perform some operations on the original instance of the snapshot to support automatic mounting of the data disk when launching new CVM instances.

What will happen when the scaling group is disabled?

After the scaling group is disabled, automatically triggered actions will not proceed, but restrictions on the scaling group are still effective.

After the scaling group is disabled, the following automatic actions will not proceed:

- Auto scaling
- Scheduled actions
- Health checks
- The desired number of instances does not match the desired capacity due to manual operations

To ensure the normal running of the business, manual scaling is still available, with the following restrictions:

- If the number of instances is less than min capacity, the removal will not be allowed.
- If the number of instances is greater than max capacity, the addition will not be allowed.
- Increasing min or max capacity manually will not trigger a scaling action.

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: Binding the CVM to a CLB
- DetachLb: Unbinding the CVM from a CLB
- Preheating: a CVM is preheating

What are the rules of CVM removal?

- Removal of a manually added CVM: the CVMs are only removed from the scaling group but not terminated. Auto Scaling will unbind these CVMs from the CLB that was automatically bound at the time of joining the scaling group, but the CLBs manually bound by the user will be not be unbound.
- Removal of the automatically added CVM: the CVMs will be terminated and unbound from the CLB.