Auto Scaling

Monitoring

Product Documentation

Copyright Notice

Trademark Notice

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

Monitoring
Monitor Alarm Metrics

Last updated : 2020-12-31 14:47:19

By using Tencent Cloud Cloud Monitoring, you can retrieve statistical data based on an ordered set of time series data (called metrics). You can use these metrics to verify whether your system is running as expected. If the thresholds are reached, scaling will be triggered.

AS monitoring metrics

AS currently supports the following metrics:

- CPU utilization
- Memory utilization
- Private network bandwidth in
- Public network bandwidth in
- Private network bandwidth out
- Public network bandwidth out

Each metric supports the following values:

- Maximum value
- Minimum value
- Average value

Metrics aggregation method

Tencent Cloud Auto Scaling monitors CVM clusters, and collects multiple monitoring metrics for multiple CVMs across multiple time periods. This data is aggregated and used to inform operations based on the policies that you configure.

It collects one sample per minute for each of CVM based on the set monitoring items. 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: A scaling group has 5 CVM instances and the defined alarm scaling policy is “Maximum/minimum/average CPU utilization of more than 50% within a 5 minutes period for 3 times

consecutively.” Auto Scaling collects the data and determines whether the policy for the alarm scaling rule is met. The steps are as follows:

1. Collects one value from each CVM per minute. In one sampling period (the current period is 5 minutes) it takes 25 CPU utilization values.

2. According to the configured maximum/minimum/average values, it determines whether the policy for the alarm scaling rule is met.

- Maximum value: If the maximum value in these 25 values exceeds the threshold (50%), this period meets the alarm scaling rule.
- Minimum value: If the minimum value in these 25 values exceeds the threshold (50%), this period meets the alarm scaling rule.
- Average value: If the average value of these 25 values exceeds the threshold (50%), this period meets the alarm scaling rule.

3. If the rule is met for 3 consecutive periods (a total of 15 minutes, with each 5-minute period considered the current period), the scaling action is triggered.

Running Status Check

Last updated : 2019-09-20 17:22:22

AS periodically performs health checks on the instances in your Auto Scaling group and identifies any instances that are unhealthy. After Auto Scaling marks an instance as unhealthy, the instance is scheduled for replacement.

Learn more about Instance Health Check >>