

Auto Scaling Startup Configuration



Tencent Cloud

Copyright Notice

©2013–2024 Tencent Cloud. All rights reserved.

The complete copyright of this document, including all text, data, images, and other content, is solely and exclusively owned by Tencent Cloud Computing (Beijing) Co., Ltd. ("Tencent Cloud"); Without prior explicit written permission from Tencent Cloud, no entity shall reproduce, modify, use, plagiarize, or disseminate the entire or partial content of this document in any form. Such actions constitute an infringement of Tencent Cloud's copyright, and Tencent Cloud will take legal measures to pursue liability under the applicable laws.

Trademark Notice



This trademark and its related service trademarks are owned by Tencent Cloud Computing (Beijing) Co., Ltd. and its affiliated companies ("Tencent Cloud"). The trademarks of third parties mentioned in this document are the property of their respective owners under the applicable laws. Without the written permission of Tencent Cloud and the relevant trademark rights owners, no entity shall use, reproduce, modify, disseminate, or copy the trademarks as mentioned above in any way. Any such actions will constitute an infringement of Tencent Cloud's and the relevant owners' trademark rights, and Tencent Cloud will take legal measures to pursue liability under the applicable laws.

Service Notice

This document provides an overview of the as-is details of Tencent Cloud's products and services in their entirety or part. The descriptions of certain products and services may be subject to adjustments from time to time.

The commercial contract concluded by you and Tencent Cloud will provide the specific types of Tencent Cloud products and services you purchase and the service standards. Unless otherwise agreed upon by both parties, Tencent Cloud does not make any explicit or implied commitments or warranties regarding the content of this document.

Contact Us

We are committed to providing personalized pre-sales consultation and technical after-sale support. Don't hesitate to contact us at 4009100100 or 95716 for any inquiries or concerns.

Contents

Startup Configuration

- Launch Configuration Overview
- Creating a Launch Configuration
- Viewing Launch Configuration List
- Modifying Launch Configurations
- Exception Detection
- Multiple-model configuration

Startup Configuration

Launch Configuration Overview

Last updated: 2024-01-17 17:53:50

The launch configuration serves as a template for automatically creating cloud servers, encompassing elements such as the image ID, instance type of the cloud server, system disk and data disk types and capacities, key pairs, security groups, and so forth.

When establishing a scaling group, it is imperative to specify a launch configuration. Once a launch configuration is created, its attributes become immutable.

Creating a Launch Configuration

Last updated: 2024-01-18 11:34:08

Scenario

Auto Scaling (AS) requires prior knowledge of the configuration to produce Cloud Virtual Machine (CVM) during expansion. You need to specify related resources in advance, such as images, data disk information, instance configurations, key pairs, security groups, and block storage devices.

It's worth noting that the launch configuration is merely a template, and machines are produced according to this template during automatic scaling. **Creating the launch configuration itself does not produce any machines, hence it is entirely free of charge, ensuring your peace of mind.**

Instructions

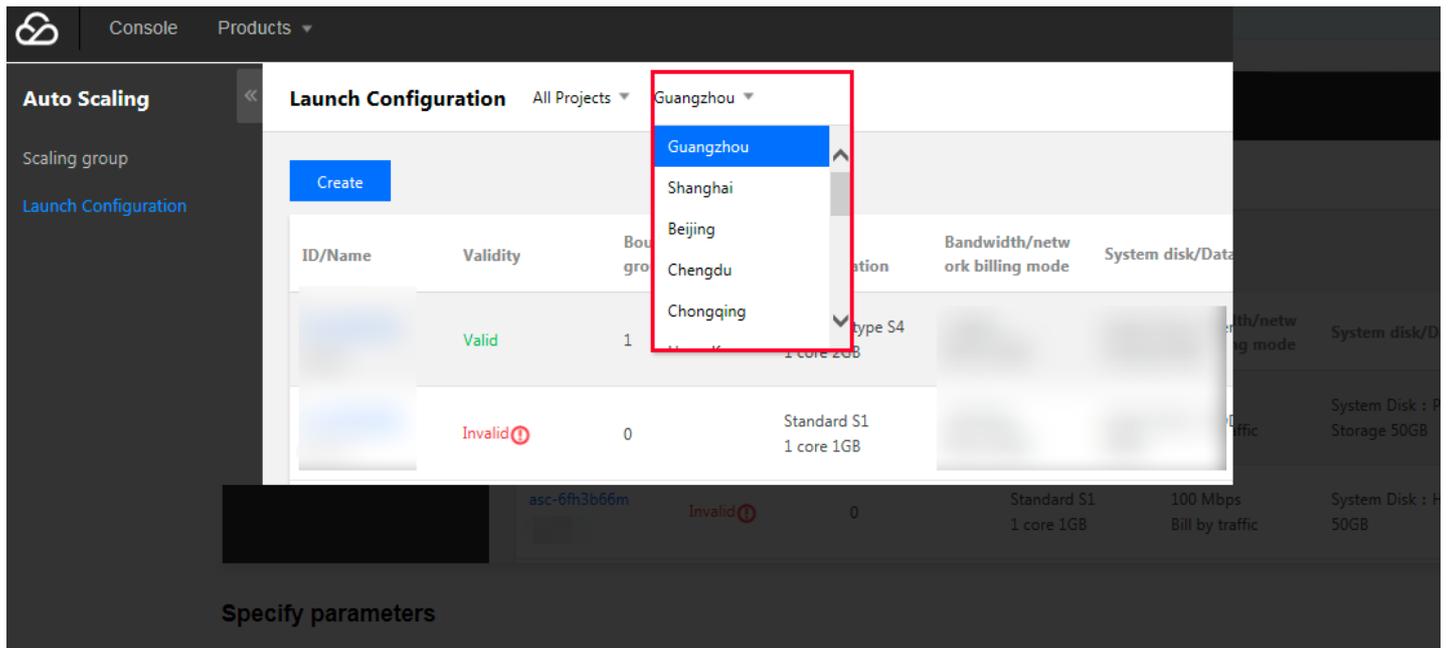
Log in to the [Auto Scaling Console](#), and click on 'Launch Configuration' in the left navigation bar.

Select region

At the top of the screen, select a region for the scaling group that meets your requirements, as shown in the figure below:

Note:

It is important to select the region where the cloud server you wish to bind to the scaling group is located, as both the launch configuration and the scaling group are region-specific. For instance, if the launch configuration is set to Guangzhou, only scaling groups in Guangzhou can be bound, and the automatically added CVM will also be in Guangzhou.



Specify parameters

1. Within the **Launch Configuration** page, click on **New** and follow the instructions to create a launch configuration, the process is identical to purchasing a cloud server.
2. Enter the **Launch Configuration Name**, for example, **Front-end Server Cluster Configuration A**.
3. Select the machine type. For example, 2-core 2G, which means a 2-core CPU with 2G of memory.
4. Select an image. You can choose a clean **Public Image**, or a **Custom Image** that has already had the business deployed.
To ensure that the machine is immediately usable upon creation, it is strongly recommended that you deploy your business application within the custom image. **The business application within the image should be set to launch alongside the operating system**, this way, machines expanded by AS can achieve automation.
5. Select the size of the system disk and data disk.
If you wish for the machine to come with data on the data disk upon activation, you can specify a data disk snapshot, this way, the machine will come with the data from the snapshot upon creation.

Note:

- Given that the machines in the scaling group are generally stateless, for convenience, it is recommended that you place the machine's inherent data into

the custom image. If the system disk is not large enough, you can submit a ticket to request a larger system disk.

- If you wish to use a data disk to store data, you need to set the data disk to mount automatically, this way, expansion can occur without manual intervention. For more details, please refer to the [specific method](#).

6. Select the bandwidth, similar to the operation when purchasing a cloud server.
7. Set the username, password, and security group.
8. Click **Finish**.
9. Create a scaling group based on this launch configuration. The launch configuration determines what machine to create during scaling, while the scaling group determines when to scale.

Viewing Launch Configuration List

Last updated: 2024-01-17 17:54:03

The launch configuration serves as a template for automatically creating cloud servers, encompassing elements such as the image ID, instance type of the cloud server, system disk and data disk types and capacities, key pairs, security groups, and so forth.

Note:

- It is imperative to specify a launch configuration when creating a scaling group.
- Once a launch configuration is created, its attributes become immutable.

Open the [Auto Scaling Console](#) and select **Launch Configuration** from the navigation bar to view the list.

- To view the details of a launch configuration, please click on the corresponding launch configuration ID.
- To delete a launch configuration, please click on **Delete** in the **Actions** column of the corresponding launch configuration entry.

Note:

A launch configuration that is bound to a scaling group cannot be deleted.

The screenshot displays the 'Launch Configuration' page in the Tencent Cloud Auto Scaling console. The page includes a navigation sidebar on the left with 'Auto Scaling' and 'Launch Configuration' options. The main content area shows a table of launch configurations with columns for ID/Name, Validity, Bound scaling group, Instance Configuration, Instance Billing Mode, Bandwidth/network billing mode, System disk/Data disk, Image, Last Modified Time, Latest Version No., and Operation. Two configurations are listed, both with a 'Valid' status. The first configuration is bound to a scaling group (ID: 1) and the second is not (ID: 0). Both use SA2.SMALL1 instances with 1 core and 1GB of memory. The first configuration was created on 2021-02-02, and the second on 2020-04-10. The page also features a 'Create' button, a search bar, and a pagination control showing 2 total items and 1 record per page.

ID/Name	Validity	Bound scaling group	Instance Configuration	Instance Billing Mode	Bandwidth/network billing mode	System disk/Data disk	Image	Last Modified Time	Latest Version No.	Operation
[Redacted]	Valid	1	SA2.SMALL1 (1 core 1GB)	Pay as you go	1 Mbps Bill by traffic	System Disk: Premium Cloud Storage 50GB	[Redacted]	2021-02-02 10:17:34	1	Delete Modify image Configure Multi-Model
[Redacted]	Valid	0	S2.SMALL1 (1 core 1GB)	Pay as you go	1 Mbps Bandwidth Package	System Disk: Premium Cloud Storage 50GB	[Redacted]	2020-04-10 14:35:03	3	Delete Modify image Configure Multi-Model

Modifying Launch Configurations

Last updated: 2024-01-17 17:54:12

Scenario

While the launch configuration itself is not editable, the scaling group can replace it. If an application upgrade or data update necessitates a change in the launch configuration, please refer to this document for guidance.

Instructions

1. Refer to [Create Launch Configuration](#) to establish a new launch configuration.
2. Navigate to the [Scaling Group](#) page, select the required scaling group ID, and proceed to its detailed page.
3. Click on **Basic Info** and select the **Edit** option in the top right corner to update the launch configuration, as illustrated below:

Basic info

Name

The name can contain up to 55 characters, including Chinese characters, English letters, numbers, underscores, hyphens and periods.

ID

Region

Min Capacity

Desired Capacity ⓘ

Current Capacity

Max Capacity

Removal policy

Activity retry policy

Launch Configuration

Supported Networks

Supported Availability Zones

Availability Zone

Guangzhou Zone 2

Guangzhou Zone 3

Guangzhou Zone 4

You can select multiple availability zones. In the case of auto-scaling up, CVMs will be created in these zones randomly to implement cross-region disaster recovery. [Suggested Settings](#)

4. Once the edits are complete, click on **Save** to finalize the changes.

Exception Detection

Last updated: 2024-01-17 17:54:25

Auto Scaling AS offers you an exceptionally considerate service: proactive anomaly detection. For instance, circumstances such as insufficient account balance or accidental deletion of images could hinder the normal production of machines required for expansion. Auto Scaling AS preemptively detects such anomalies and issues warnings. You need not wait until the scaling operation fails to identify risks, thus nipping potential losses in the bud.

You can ascertain the cause of the anomaly through the following methods:

Directly inspect the launch configuration list. If the **Validity** is displayed as **Invalid**, it indicates that your launch configuration has been mishandled, rendering it unusable. Hovering over it will reveal the reason for invalidity.

ID/Name	Validity	Bound scaling group	Instance configuration	Instance billing mode	Bandwidth /Network billing	System disk/Data disk	Image	Last modified	Latest version	Operation
	Valid	3	SN3ne.SMALL2 (1 core 2GB)	Pay-as-you-go	0 Mbps Bill by traffic	System disk: Premium Cloud Disk 20GB			1	Delete Modify image Configure mul
	Invalid 	1	SN3ne.MEDIUM4 (2 core 4GB)	Pay-as-you-go	1 Mbps Bill by traffic	System disk: Premium Cloud Disk 50GB			6	Delete Modify image Configure mul

Multiple-model configuration

Last updated: 2024-01-17 17:54:37

Operational Overview

The availability of instance types varies across each availability zone. You may encounter issues such as the selected model being sold out in the current availability zone or related configuration problems. Configuring multiple instance types can effectively address these issues.

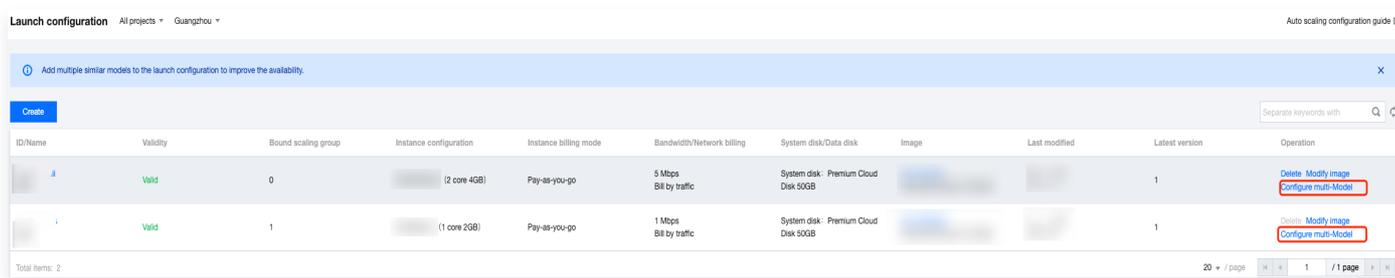
The multi-instance configuration feature possesses the following characteristics:

- This effectively mitigates the risk of scaling failure brought about by sell-outs or instance invocation failures.
- We recommend selecting a variety of models that are similar to your preferred model in terms of configuration, price, and performance. In the event that your preferred model is sold out, the system can intelligently select other models with abundant inventory for you.

Instructions

Multiple-model configuration

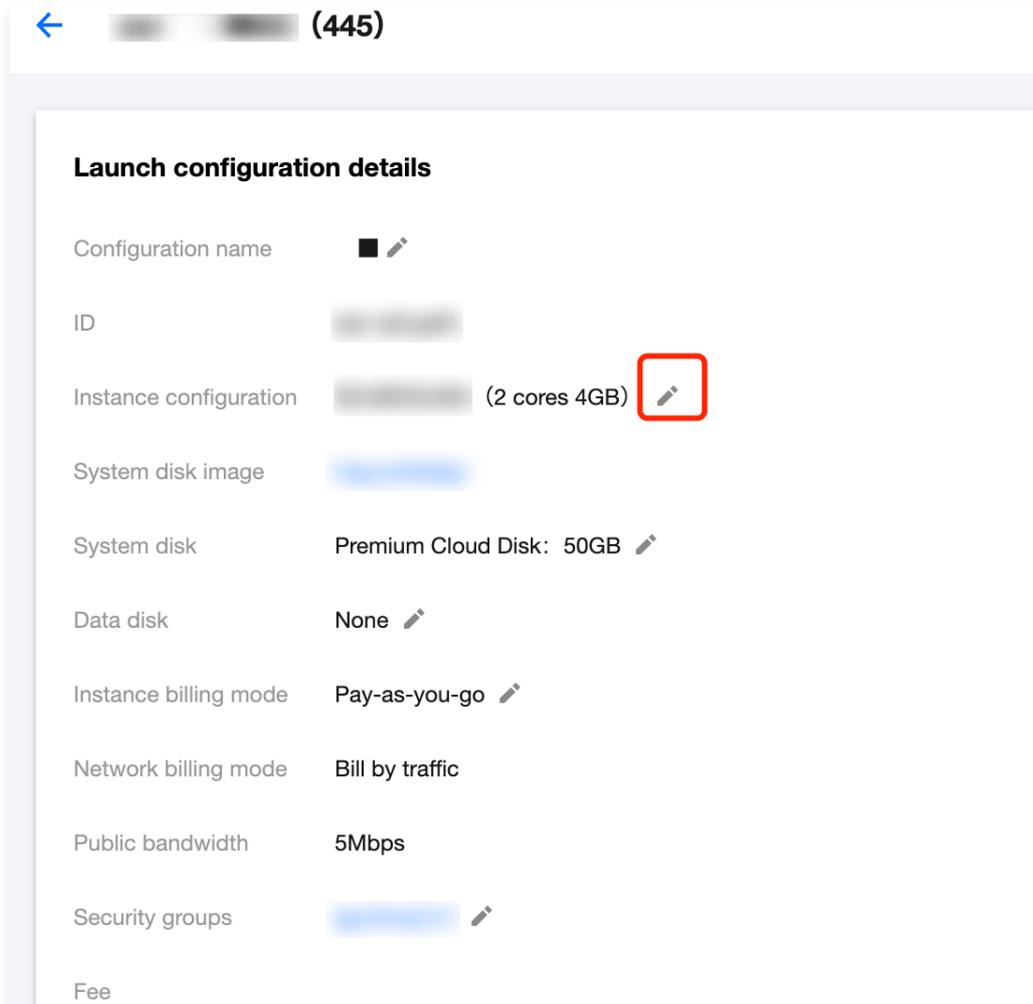
1. Log in to the [Auto Scaling Console](#), and click on **Launch Configuration** in the left navigation bar.
2. On the **Launch Configuration** page, you can access the multiple-model configuration page in two ways:
 - Click on **Configure Multiple Models** on the right side of the row where the multi-model launch configuration needs to be set, as shown in the following figure:



The screenshot shows the 'Launch configuration' page in the Auto Scaling Console. At the top, there is a navigation bar with 'All projects' and 'Guangzhou'. Below it, a blue banner contains the text 'Add multiple similar models to the launch configuration to improve the availability.' and a close button. A 'Create' button is visible on the left. The main content is a table with columns: ID/Name, Validity, Bound scaling group, Instance configuration, Instance billing mode, Bandwidth/Network billing, System disk/Data disk, Image, Last modified, Latest version, and Operation. Two rows are visible, both with 'Valid' status. The 'Operation' column for both rows contains 'Delete' and 'Modify image' buttons, with 'Configure multi-Model' buttons highlighted in red boxes. At the bottom right, there is a pagination control showing '20 / page' and '1 / 1 page'.

ID/Name	Validity	Bound scaling group	Instance configuration	Instance billing mode	Bandwidth/Network billing	System disk/Data disk	Image	Last modified	Latest version	Operation
	Valid	0	(2 core 4GB)	Pay-as-you-go	5 Mbps Bill by traffic	System disk- Premium Cloud Disk 50GB			1	Delete Modify image Configure multi-Model
	Valid	1	(1 core 2GB)	Pay-as-you-go	1 Mbps Bill by traffic	System disk- Premium Cloud Disk 50GB			1	Delete Modify image Configure multi-Model

- Click on the ID of the launch configuration that requires multiple-model configuration, and select  on the **Launch Configuration Details** page, as illustrated in the following figure:



3. On the **Multiple-model Configuration** page, select any machine type you need and click **Confirm**, as depicted in the following figure:

Multi-model configuration ✕

i Configuring multi-model can help reduce the risk of scale-out failure due to resource sold-out or model not matched. Available instance models vary for different availability zones. We recommend choosing multiple models with different price and similar performance. When your preferred model is sold out, AS will select from your alternative models automatically. The 1-core models are in short supply. It is recommended to choose a 2C2G model, which is priced the same as the 1C2G model.

Your current preferred model configuration is (2-core 4 GB). You can select (or cancel) the model configuration and reset the model priority:

Display only recommended configurations

Select model

Please enter the model configuration, such as S5.SMALL2 🔍

<input type="checkbox"/>	Model	CPU MEM	Supported ...	Price
<input checked="" type="checkbox"/>	S6.MEDIUM4	2C4G	Guangzhou Zone 6, Guangzhou Zone 7	0.06USD/hour
<input type="checkbox"/>	S6.MEDIUM2	2C2G	Guangzhou Zone 6, Guangzhou Zone 7	0.04USD/hour
<input type="checkbox"/>	S6.MEDIUM8	2C8G	Guangzhou Zone 6, Guangzhou Zone 7	0.08USD/hour

Selected (1)

Model	CPU MEM	Supported...	Price
⋮ S6.MEDIUM4	2C4G	Guangzhou Zone 6, Guangzhou Zone 7	0.06USD/hour

Viewing Multiple-Model Configuration

Upon completion of the configuration, you can view the configured multiple models and their associated costs on the details page of this launch configuration, as illustrated in the following figure:

Launch configuration details

Configuration name	
ID	asc-q2ug4lii
Instance configuration	<div style="border: 2px solid red; padding: 5px;"><p>S6.MEDIUM4 (2 cores 4GB)</p><p>S6.MEDIUM2 (2 cores 2GB) </p><p>S6.MEDIUM8 (2 cores 8GB)</p></div>
System disk image	img-evitcbqz TencentOS Server 2.6 (Final)
System disk	Premium Cloud Disk: 50GB
Data disk	None
Instance billing mode	Pay-as-you-go
Network billing mode	Bill by traffic
Public bandwidth	5Mbps
Security groups	sg-krrcq1c1
Fee	0.2 USD/hour (configuration) 0.2 USD/GB (network)

Modifying Multiple-Model Configuration

If the completed multiple-model configuration does not meet your expectations, you may modify it following the steps outlined below.

1. Refer to [Multiple-model Configuration](#) to access the **Multiple-model Configuration** page.
2. After modifying the original model configuration, click **Confirm** to complete the changes.