Auto Scaling

FAQ

FAQ

Environment configuration

Environment configuration

How can I ensure the configuration environments are consistent with Auto Scaling ECS instances? Alternatively, how can I directly add a previous system environment to the Auto Scaling instances?

When creating scaling configuration for ECS instances, you can use an ECS custom image template to create instances. If you need to sync internal system data (such as a previous system environment) when the ECS instances are running, we recommend that you install a custom rsync program.

In the created instances, after a restart, why is 127.0.0.1 added after /etc/hosts are cleared?

When your content is added after /etc/hosts in an image, and the custom image generated by this image is used to create an instance, its configuration will be restored to the system default settings. This means the added content will be cleared. If these settings need to be retained, add the script code in rc.local. Then, check if the information is in /etc/hosts. If not, you must add the script code again.

My Auto Scaling service is set up to automatically create instances, however, there is no fixed quantity. How can I ensure that my instances are scaled normally when using images in the image market?

If you need to scale to N instances that use the same image, you must buy N images from the image market in advance.

Can I buy images from the image market in batches?

Batch purchasing of images is not supported.

If a previously used image from the image market no longer exists, how can I ensure that the scaling group instances set up can be scaled normally?

We recommend that you select a suitable replacement image from the image market to ensure your scaling group can scale normally.

Can I use a single product code to get images from different regions?

Yes. However, you must ensure that your desired regions support the image.

I have bought 100 images with the same product code. Can I use these images in all regions?

Images in the image market have region attributes. You must ensure the purchased images are supported by your desired regions.

Password and logon

Password and login troubleshooting

When Auto Scaling automatically creates instances, how do I view their passwords and subsequently log in to these instances?

Instances automatically created by Auto Scaling do not have the same password. In a Linux environment, we recommend that you set a public/private key certificate for SSH login without password.

If you do not want to set a public/private key certificate for SSH login without password, you

must reset the password on the console, and then restart the instance to apply the new password, before you can log in.

Why are the passwords of instances created by Auto Scaling different from the password for my custom image?

Created ECS instances do not have the same password as the custom image. To ensure password security, we recommend that you set a public/private key certificate for SSH password-free logon.

If you do not want to set a public/private key certificate for SSH login without password, you must reset the password on the console, and then restart the instance to apply the new password, before you can log in.

When using a custom image to generate Linux system instances, can I manage the instances through SSH login without password?

You can set a public/private key certificate for SSH login without password as follows:

Establish a public key and private key in the custom image’ s ECS server-end instance.

Copy the ECS instance idc.pub to the client.

Delete the public key from the server-end.

Modify the SSH configuration file on the server-end.

Configure the client software.

Using SecureCRT configuration as an example:

Select the corresponding remote connection information.

Right-click on the **Attribute** option and select **SSH2**.

Clear the **Password** option and select **PublicKey**.

Click the **Attribute** button on the right side and select **Use Session Public Key Settings**

Select **Use ID or Certificate File**, and **idc.pub** (the public and private key files you previously copied from the server).

Scaling configuration rules FAQ

What information should I provide for Auto Scaling troubleshooting?

When you open a ticket, we recommend that you provide your Auto Scaling activity ID (ScalingActivityId) and relevant logs to facilitate troubleshooting.

Can I add ECS instances that I have already created to a scaling group?

Yes. However, the ECS instances:

Must be in the same region of the scaling group. For more information, see Regions and zones.

Must be in the **Running** status. For more information, see *ECS* Instance life cycle.

Cannot exist in more than one scaling group.

Can I add existing subscription instances to scaling groups?

Yes, you can. Auto Scaling automatically creates Pay-As-You-Go or spot instances by default, and you can also add your existing Subscription or Pay-As-You-Go instances to a scaling group.

Can I add an ECS instance to more than one scaling group?

No, you cannot. This feature is not currently supported.

Can I create 8vCPU and 16vCPU ECS instances in a scaling group?

Yes, you can. Open a ticket to use more ECS instance types when you create ECS instances.

Is Auto Scaling vertical scalable?

No, it is not. Auto Scaling does not support automatically upgrade or downgrade the vCPU, memory, or bandwidth of an ECS instance.

Can I create periodic tasks in Auto Scaling?

Yes, you can. For more information, see Create a scheduled task.

When I remove an ECS instance from a scaling group and release the instance, can I save the ECS instance data?

No, you cannot. Therefore, do not establish application status information (for example, session) or related data (such as databases and logs) in an Auto Scaling ECS instances. We recommend that you save the status information to an independent state server (for example, ECS), database (for example, RDS), or standardized log storage (for example, Log Service).

If I disable a scaling group, are instances created by Auto Scaling released?

No. After you disable a scaling group (DisableScalingGroup), instances created by Auto Scaling are not automatically released.

Can I add different ECS instance types in an Auto Scaling groups?

No, you cannot. However, each scaling group can be set with a different configuration type.

Can I increase the maximum number of instances in the Auto Scaling service?

Yes, you can. The default maximum number of ECS instance in a scaling group is 1,000. Open a ticket for a higher instance quota.

Does Auto Scaling automatically include or exclude a new or removed ECS instances into or from the IP address whitelists of the configured RDS or Memcache?

Auto Scaling automatically includes a new or removed ECS instances into or from the IP address whitelists of a RDS. However, Memcache whitelists are not supported.

How can I make sure that manually added ECS instances are not removed from the scaling group?

For automatically created ECS instances, supposing that you want to retain the specified **100** ECS instances in a scaling group, and pay attention to the following when you Create a scaling configuration(CreateScalingConfiguration):

Set the minimum number of instances to greater than **100**.

Set the first Removal Policy to **The instances with the oldest configuration**.

For manually created ECS instances, do not stop the specified ECS instance. Because the manually created and added ECS instances are not released if they are removed from a scaling group.

Note:Since manually added ECS instances were not created by a scaling group, therefore, Auto Scaling removes the automatically created ECS instances from the scaling group. Manually added ECS instances are removed only after all the automatically created ECS instances have been removed.

Auto Scaling, Server Load Balancer, and RDS FAQ

- After Auto Scaling creates an ECS instance, will the new instance be automatically added to a Sever Load Balancer instance?
- When a scaling group is added in Auto Scaling, can I bind multiple Sever Load Balancer instances to the group?
- When Auto Scaling creates an ECS instance, can the instance be added to multiple Server Load Balancer instances?
- Can I modify the weights of ECS instances added to an Auto Scaling group Server Load Balancer Instance?
- I have a public network Server Load Balancer instance. If I create a scaling configuration, will its ECS instances need public bandwidth?
- Do I need to use Server Load Balancer, CloudMonitor, and RDS in combination with Auto Scaling?

After Auto Scaling creates an ECS instance, will the new instance be automatically added to a Sever Load Balancer instance?

If a Server Load Balancer instance is specified in a scaling group, the scaling group will automatically add the ECS instances in the group to the specified Server Load Balancer instance.

When a scaling group is added in Auto Scaling, can I bind multiple Sever Load Balancer instances to the group?

By default, you can only bind five Server Load Balancer instance to each scaling group. To bind multiple Server Load Balancer instances, open a ticket to apply for a higher quota to Alibaba Cloud Technical Support.

When Auto Scaling creates an ECS instance, can the instance be added to multiple Server Load Balancer instances?

Yes. However, you must open a ticket to Alibaba Cloud Technical Support to apply for multiple Server Load Balancer instances to bind to a specified ECS instance.

Can I modify the weights of ECS instances added to an Auto Scaling group Server Load Balancer Instance?

Yes. By default, ECS instances under Auto Scaling Sever Load Balancer instances have a weight of 50. This is suitable for most scenarios, as backend ECS instances of Auto Scaling groups normally carry the same services and are the same type. The Server Load Balancer also distributes traffic based on the weight ratio, not the actual number. This means that, if you have two backend ECS instances weighted 50 and 50 (with a ratio of 1:1), this is the same as if they were weighted 100 and 100.

I have a public network Server Load Balancer instance. If I create a scaling configuration, will its ECS instances need public bandwidth?

When a scaling configuration is created, you do not have to allocate public bandwidth to its ECS instances. However, we recommend that you set at least 1 Mbit/s of ECS bandwidth when creating a scaling configuration, for easier ECS instance management.

Do I need to use Server Load Balancer, CloudMonitor, and RDS in combination with Auto Scaling?

Auto Scaling is an open elastic scaling platform, and can independently scale up or down ECS instances. It can be deployed either separately or in combination with the Server Load Balancer and ApsaraDB for RDS. Auto Scaling allows CloudMonitor to trigger scaling up or scaling down actions for ECS instances.

Monitoring and automation

Monitoring and automation

How does Auto Scaling determine if its ECS instances are available?

If the Server Load Balancer is available in the expected Auto Scaling group, it will check that the ports of the backend ECS instances are functional before forwarding requests to the ECS instances.

What are the triggering conditions for Auto Scaling alarms?

Monitoring alarms in Auto Scaling are triggered based on the CPU load, memory usage, average system load, and Internet and intranet inbound and outbound traffic. These are used to automatically increase or decrease the number of ECS instances.

Can Auto Scaling support dynamic scaling based on custom alarms in CloudMonitor?

No. Dynamic scaling based on custom monitoring settings is not supported.

How can I automate the deployment of the ECS applications

created in a scaling group?

To automatically install or update a program, or automatically load code after an ECS instance is automatically created in a scaling group, you must store an execution script in a custom image and set up a command to automatically run this script upon operating system startup.

Take CentOS as an example, you can create the following shell test script:

```
#!/bin/sh
# chkconfig: 6 10 90
# description: Test Service
echo "hello world!"
```

In the preceding output, 6 is the default start level. There are a total of 7 levels ranging from 0-6.

- Level 0: Shutdown
- Level 1: Single user mode
- Level 2: Multiuser command line mode with no network connection
- Level 3: Multiuser command line mode with network connection
- Level 4: Unavailable
- Level 5: Multiuser mode with graphic interface
- Level 6: Restart
- 10 is the start priority and 90 is the stop priority. The priority range is 0-100. The higher the number, the lower the priority.

Put the test file in the **/etc/rc.d/init.d/** directory and run chkconfig --level 6 test on. This test script will run each time the system starts up.

The following example shows how to use a script to install Phpwind. Put the Phpwind installer in the script for execution (you will need to enter the database password). An example output is as follows:

```
cd /tmp

echo "phpwind"

yum install -y \
unzip \
wget \
httpd \
php \
php-fpm \
php-mysql \
php-mbstring \
php-xml \
php-gd \
php-pear \
php-devel
chkconfig php-fpm on \
&& chkconfig httpd on
```

```
wget http://pwfiles.oss-cn-hangzhou.aliyuncs.com/com/soft/phpwind_v9.0_utf8.zip \
&& unzip -d pw phpwind_v9.0_utf8.zip \
&& mv pw/phpwind_v9.0_utf8/upload/* /var/www/html \
&& wget http://ess.oss-cn-hangzhou.aliyuncs.com/ossupload_utf8.zip -O ossupload_utf8.zip \
&& unzip -d ossupload ossupload_utf8.zip \
&& /bin/cp -rf ossupload/ossupload_utf8/* /var/www/html/src/extensions/ \
&& chown -R apache:apache /var/www/html
service httpd start && service php-fpm start
echo "Install CloudMonitor"
wget http://update2.aegis.aliyun.com/download/quartz_install.sh
chmod +x quartz_install.sh
bash quartz_install.sh

echo "Installation complete"
```