Alibaba Cloud
Auto Scaling

Best practices

Issue: 20181115

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade

secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion , or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos , marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).

6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
⊖	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	⊖ **Danger:** Resetting will result in the loss of user configuration data.
⚠	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	⚠ **Warning:** Restarting will cause business interruption. About 10 minutes are required to restore business.
📋	This indicates warning information, supplementary instructions, and other content that the user must understand.	📋 **Note:** Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	📋 **Note:** You can use **Ctrl** + **A** to select all files.
>	Multi-level menu cascade.	**Settings** > **Network** > **Set network type**
Bold	It is used for buttons, menus, page names, and other UI elements.	Click **OK**.
`Courier font`	It is used for commands.	Run the `cd /d C:/windows` command to enter the Windows system folder.
Italics	It is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	It indicates that it is a optional value, and only one item can be selected.	`ipconfig` *[-all|-t]*
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`swich` *{stand	slave}*

Contents

1 API quick start

1.1 Process introduction

This topic explains how to use APIs to create and configure overall scaling solutions, including scheduled, dynamic, customized, and fixed quantity scaling.

Follow the steps shown in the following figure. The first three steps are used to create a simple scaling solution:

1. Create a scaling group. Configure the minimum (Min Size) and maximum (Max Size) number of ECS instances for scaling, and select a Server Load Balancer instance and an RDS instance.

2. Create scaling configuration. Specify the specifications of ECS instances used for Auto Scaling , such as Image ID and Instance Type.

3. Enable the scaling group according to the scaling configuration created in Step 2.

4. Create a scaling rule. For example, **adding N (number) ECS instances**.

5. Create a scheduled task. For example, triggering the scaling rule at 13:00.

6. Create an alarm task (CloudMonitor API PutAlarmRule). For example, adding 1 ECS instance when the CPU usage is equal to or greater than 80%.

1.2 Create a scaling group

This topic gives you an example of how to make the CreateScalingGroup API request, and specifying MinSize and MaxSize.

```
http://ess.aliyuncs.com/?Action=CreateScalingGroup
&RegionId=cn-qingdao
&MaxSize=20
&MinSize=2
&LoadBalancerId=147b46d767c-cn-qingdao-cm5-a01
&DBInstanceId. 1=rdszzzyyunybaeu
&DBInstanceId. 2=rdsia3u3yia3u3y
&<Public Request Parameters>
```

Response example

```
<CreateScalingGroupResponse>
    <ScalingGroupId>dP8VqSd9ENXPc0ciVmbcrBT1</ScalingGroupId>
    <RequestId>536E9CAD-DB30-4647-AC87-AA5CC38C5382</RequestId>
```

```
</CreateScalingGroupResponse>
```

1.3 Create a scaling configuration

You must specify the attributes of the ECS instances to be scaled, such as the ImageID and

InstanceType. You must specify the ScalingGroupId returned in step 1.

```
http://ess.aliyuncs.com/?Action=CreateScalingConfiguration
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbcrBT1
&SecurityGroupId=sg-280ih3w4b
&ImageId=centos6u5_64_20G_aliaegis_20140703.vhd
&InstanceType=ecs.t1.xsmall
&<common request parameters>
```

Response example

```
<CreateScalingConfigurationResponse>
    <ScalingConfigurationId>eOs27Kb0oXvQcUYjEGelJqUy</ScalingCon
figurationId>
    <RequestId>5CC0AD41-08ED-4559-A683-6F56355FE068</RequestId>
</CreateScalingConfigurationResponse>
```

1.4 Enable a scaling group

This topic gives you an example of how to make the EnableScalingGroup API request according

to the scaling configuration you specified, and add an existing ECS instance.

```
http://ess.aliyuncs.com/?Action=EnableScalingGroup
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbcrBT1
&ActiveScalingConfigurationId=eOs27Kb0oXvQcUYjEGelJqUy
&InstanceId. 1=i-283vvyytn
&<Public Request Parameters>
```

Response example

```
< EnableScalingGroupResponse>
    <RequestId>6469DCD0-13AC-487E-85A0-CE4922908FDE</RequestId>
</ EnableScalingGroupResponse>
```

1.5 Create a scaling rule

This topic is an example of API request `CreateScalingRule`(create a scaling rule). You must

specify the `ScalingGroupId`.

```
http://ess.aliyuncs.com/?Action=CreateScalingRule
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbcrBT1
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=1
```

```
&<common request parameters>
```

Response example

```
<CreateScalingRuleResponse>
<ScalingRuleAri>
ari:acs:ess:cn-qingdao:1344371:scalingrule/eMKWG8SRNb9dBLAjweNI1Ik
</ScalingRuleAri>
    <ScalingRuleId>eMKWG8SRNb9dBLAjweNI1Ik</ScalingRuleId>
    <RequestId>570C84F4-A434-488A-AFA1-1E3213682B33</RequestId>
</CreateScalingRuleResponse>
```

1.6 Create a scheduled task

This topic is an example of API request `CreateScheduledTask` (create a scheduled task). You

must specify the `ScalingRuleAri`.

```
http://ess.aliyuncs.com/?Action=CreateScheduledTask
&RegionId=cn-qingdao
&LaunchTime=2014-08-17T12:00Z
&RecurrenceType=Daily
&RecurrenceValue=1
&RecurrenceEndTime=2014-09-17T16:55Z
&ScheduledAction=ari:acs:ess:cn-qingdao:1344371:scalingrule/eMKWG8SRNb
9dBLAjweNI1Ik
&<common request parameters>
```

Response example

```
<CreateScheduledTaskResponse>
    <ScheduledTaskId>edRtShc57WGXdt8TlPbrjsnV</ScheduledTaskId>
    <RequestId>0F02D931-2B12-44D7-A0E9-39925C13D15E</RequestId>
</CreateScheduledTaskResponse>
```