

Auto Scaling

Quick Start

Issue 01
Date 2021-12-31



Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2023. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Cloud Computing Technologies Co., Ltd.

Trademarks and Permissions



HUAWEI and other Huawei trademarks are the property of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei Cloud and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Contents

1 Scaling Your Website Dynamically.....	1
2 Scaling Your Website on a Schedule.....	2
3 Wizard-based Process of Using AS.....	3
4 Creating an AS Group Quickly.....	4

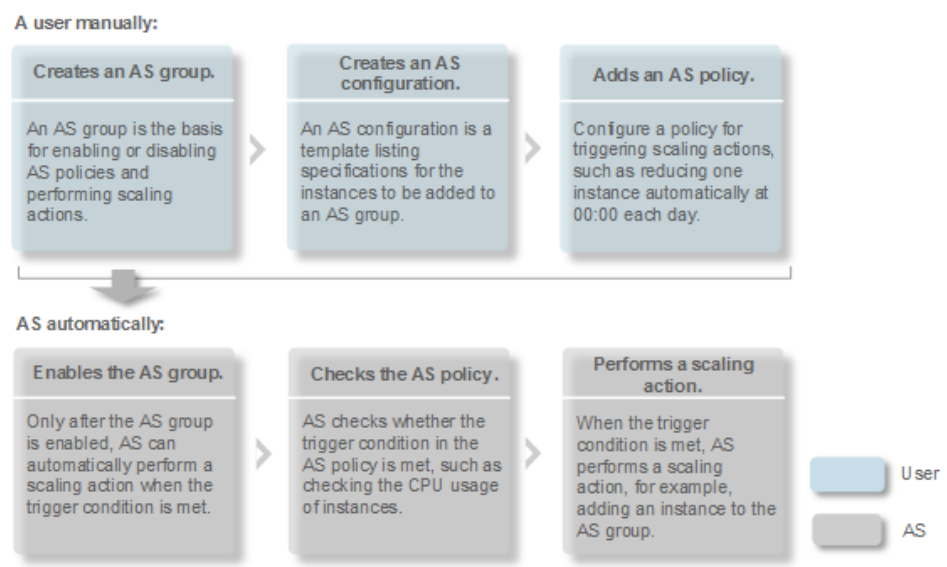
1 Scaling Your Website Dynamically

2 Scaling Your Website on a Schedule

3 Wizard-based Process of Using AS

Figure 3-1 illustrates the wizard-based process of using AS.

Figure 3-1 Wizard-based process of using AS



4 Creating an AS Group Quickly

If you are using AS for the first time, following the wizard-based process is an easy way to create an AS group, AS configuration, and AS policy.

Prerequisites

- You have created the required VPCs, subnets, security groups, and load balancers.
- You have obtained the key pairs for logging in to the instances added by a scaling action if key authentication is used.

Procedure

1. Log in to the management console.
2. Under **Compute**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.
3. Click **Create AS Group**.
4. Set basic information about the AS group, such as **Name**, **Max. Instances**, **Min. Instances**, and **Expected Instances**. [Table 4-1](#) lists the parameters.

Table 4-1 AS group parameters

Parameter	Description	Example Value
Region	A region is where the AS group is deployed. Resources in different regions cannot communicate with each other over internal networks. For lower network latency and faster access to your resources, select the region nearest to your target users.	N/A

Parameter	Description	Example Value
AZ	<p>An AZ is a physical location where resources use independent power supply and networks. AZs are physically isolated but interconnected through an internal network.</p> <ul style="list-style-type: none"> • If you require high availability, buy servers in different AZs. • If you require low network latency, buy servers in the same AZ. 	N/A
Multi-AZ Scaling Policy	<p>This parameter can be set to Load-balanced or Sequenced.</p> <ul style="list-style-type: none"> • Load-balanced: When scaling out an AS group, the system preferentially distributes ECS instances evenly among AZs used by the AS group. If it fails in the target AZ, it automatically selects another AZ based on the sequenced policy. • Sequenced: When expanding ECSs in an AS group, the system selects the target AZ based on the order in which AZs are selected. <p>NOTE This parameter needs to be configured when two or more AZs are selected.</p>	Load-balanced
Name	<p>Specifies the name of the AS group to be created.</p> <p>The name contains 1 to 64 characters and consists of only letters, digits, underscores (_), and hyphens (-).</p>	N/A
Max. Instances	Specifies the maximum number of ECS instances in an AS group.	1
Expected Instances	<p>Specifies the expected number of ECS instances in an AS group.</p> <p>After an AS group is created, you can change this value, which will trigger a scaling action.</p>	0
Min. Instances	Specifies the minimum number of ECS instances in an AS group.	0
VPC	<p>Provides a network for your ECS instances.</p> <p>All ECS instances in the AS group are deployed in this VPC.</p>	N/A

Parameter	Description	Example Value
Subnet	You can select up to five subnets. The AS group automatically binds all NICs to the created ECSs. The first subnet is used by the primary NIC of an ECS instance by default, and other subnets are used by extension NICs of the instance.	N/A
Load Balancing	<p>This parameter is optional. A load balancer automatically distributes traffic across all instances in an AS group to balance their service load. It improves the fault tolerance of your applications and expands application service capabilities.</p> <p>NOTE</p> <ul style="list-style-type: none">• Up to six load balancers can be added to an AS group.• After multiple load balancers are added to an AS group, multiple services can be concurrently listened to, thereby improving scalability. If ELB health check is selected for Health Check Method, when any one of the listeners detects that an instance becomes unhealthy, AS will replace the faulty instance with a functional one. <p>If you select Elastic load balancer, configure the following parameters:</p> <ul style="list-style-type: none">• Load Balancer• Backend ECS Group• Backend Port: specifies the port on which a backend ECS listens for traffic.• Weight: determines the portion of requests a backend ECS processes compared to other backend ECSs added to the same listener. <p>For more information about load balancing, see Elastic Load Balance User Guide.</p>	N/A

Parameter	Description	Example Value
Instance Removal Policy	<p>Specifies the priority for removing instances from an AS group. If specified conditions are met, scaling actions are triggered to remove instances. AS supports the following instance removal policies:</p> <ul style="list-style-type: none">● Oldest instance created from oldest AS configuration: The oldest instance created from the oldest configuration is removed from the AS group first.● Newest instance created from oldest AS configuration: The newest instance created from the oldest configuration is removed from the AS group first.● Oldest instance: The oldest instance is removed from the AS group first.● Newest instance: The newest instance is removed from the AS group first. <p>NOTE</p> <ul style="list-style-type: none">● Removing instances will preferentially ensure that the remaining instances are load balanced in AZs.● Manually added ECS instances are the last to be removed. If AS does remove a manually added instance, it only removes the instance from the AS group. It does not delete the instance. If multiple manually added instances must be removed, AS preferentially removes the earliest-added instance first.	N/A
EIP	<p>If EIP has been selected in an AS configuration for an AS group, an EIP is automatically bound to the ECS instance added to the AS group. If you select Release, the EIP bound to an instance is released when the instance is removed from the AS group. Otherwise, the system unbinds the EIP from the instance, but does not release it when the instance is removed from the AS group.</p>	N/A

Parameter	Description	Example Value
Health Check Method	<p>When a health check detects an unhealthy ECS instance, AS replaces the faulty with a new one. There are two health check methods:</p> <ul style="list-style-type: none">• ECS health check: checks ECS instance health status. If an instance is stopped or deleted, it is considered to be unhealthy. This method is selected by default. Using this method, the AS group periodically evaluates the running status of each instance based on the health check results. If the health check results show that an instance is unhealthy, AS removes the instance from the AS group.• ELB health check: determines ECS instance running status using a load balancing listener. This health check method is only available if the AS group uses a load balancing listener. If a load balancing listener detects that an ECS instance is unhealthy, AS removes the instance from the AS group.	N/A
Health Check Interval	<p>Specifies the length of time between health checks. You can set a health check interval, such as 10 seconds, 1 minute, 5 minutes, 15 minutes, 1 hour, or 3 hours, based on the site requirements.</p>	5 minutes
Enterprise Project	<p>Specifies the enterprise project to which the AS group belongs. If an enterprise project is configured for an AS group, ECSs created in this AS group also belong to this enterprise project. If you do not specify an enterprise project, the default enterprise project will be used.</p> <p>NOTE</p> <ul style="list-style-type: none">• Value default indicates the default enterprise project. Resources that are not allocated to any enterprise projects under your account are displayed in the default enterprise project.• Enterprise project is an upgraded version of IAM. It allocates and manages resources of different projects.	N/A
Advanced Settings	<p>Configure notifications. You can select Do not configure or Configure now.</p>	N/A

Parameter	Description	Example Value
Notification	<p>Results of scaling actions are sent to you based on the functions provided by the Simple Message Notification (SMN) service.</p> <ul style="list-style-type: none">• Notification Conditions: When at least one of the following conditions is met, SMN sends a notification to you:<ul style="list-style-type: none">- Instance creation succeeds- Instance removal succeeds- Errors occur in an AS group- Instance creation fails- Instance removal fails• Send Notification To: Select an existing topic. For details about how to create a topic, see Simple Message Notification User Guide.	N/A
Tag	<p>If you have many resources of the same type, you can use tags to manage your resources. You can identify specified resources quickly using the tags allocated to them.</p> <p>Each tag contains a key and a value. You can specify the key and value for each tag.</p> <ul style="list-style-type: none">• Key<ul style="list-style-type: none">- The key must be specified.- The key must be unique to the AS group.- The key can include up to 36 characters. It cannot contain non-printable ASCII characters (0-31) or the following characters: =*<>\\, /• Value<ul style="list-style-type: none">- The value is optional.- A key can have only one value.- The value can include up to 43 characters. It cannot contain non-printable ASCII characters (0-31) or the following characters: =*<>\\, /	N/A

5. Click **Next**.
6. On the displayed page, you can use an existing AS configuration or create an AS configuration.

7. Click **Next**.
8. (Optional) Add an AS policy to an AS group.

On the displayed page, click **Add AS Policy**.

Configure the required parameters, such as the **Policy Type**, **Scaling Action**, and **Cooldown Period**.

 **NOTE**

- If a scaling action is triggered by an AS policy, the cooldown period is whatever configured for that AS policy.
 - If a scaling action is triggered by manually changing the expected number of instances or by other actions, the cooldown period is whatever configured for the AS group. The default cooldown period is 300 seconds.
9. Click **Create Now**.
 10. Check the AS group, AS configuration, and AS policy information. Click **Submit**.
 11. Confirm the creation result and go back to the **AS Groups** page as prompted. After the AS group is created, its status changes to **Enabled**.