Auto Scaling

Best Practice

Issue	01
Date	2025-07-16

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions


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 Setting Up an Automatically Scalable Discuz! Forum

Application Scenarios

Discuz! is one of the most mature and widely used forum software programs. Website traffic reaches its top in peak hours. If a forum website is deployed on multiple servers to cope with peak hours, some servers must be sitting idle at other times. This increases the costs and leads to a waste of resources.

AS helps you address the preceding problems. After you use AS on the servers for deploying a forum website, AS maintains the proper running of the website and reduces costs by automatically adjusting the number of servers based on the predefined scaling policy. This section describes how to use AS to automatically increase or reduce the number of ECSs for a web service, for example, Discuz!.

Solution Introduction

Table 1-1 lists the steps required to set up an automatically scalable Discuz! forum website. This section focuses on how to implement AS. The increase or decrease of website traffic causes the rise or drop of the ECS CPU usage. You can configure two CPU usage alarm policies, one for increasing one ECS when the CPU usage exceeds 70% and the other for reducing one ECS when the CPU usage is less than 30%. In this way, there will remain a proper number of ECSs for the forum website.

Table 1-1 Setting up a Discuz! forum

Step	Category	Substep	Description
Setting up a website	Applying for resources	Creating a VPC	Create a VPC, for example, **vpc-DISCUZ**, that provides network services for the ECS on which the website is deployed.
		Purchasing an EIP	Purchase an EIP that allows ECSs to access the Internet.

Step	Category	Substep	Description
		Creating a security group and adding security group rules	To ensure network security, create a security group, for example, **sg-DISCUZ**, to control network access.
		Purchasing ECSs	Buy two ECSs, for example, discuz01 for deploying the forum database and discuz02 for deploying the forum. Bind the purchased EIP to the ECS discuz01 when purchasing discuz01.
	Configuring the ECSs	Installing the database on discuz01	Install an RDS for MySQL DB instance on discuz01, start the database, and configure it to automatically start upon ECS startup.
		Deploying the website code on discuz02	Unbind the EIP from discuz01, bind it to discuz02, and deploy a web environment and website code on discuz02.
	Configuring features	Unbinding the EIP	To save EIP resources, release the EIP bound to discuz02 before using the load balancing service.
		Configuring ELB	Purchase an enhanced load balance listener, for example, **elb-DISCUZ**, to balance the website traffic in an AS group.
		Creating an image	To ensure that the ECSs to be added to the AS group automatically deploy the web environment and website code, create an image, for example, **discuz_centos6.5 (40 GB)** based on discuz02. The image is used as a private one for creating an AS configuration.
Configuring AS	-	Creating an AS configuration	An AS configuration is an ECS template in the AS group, specifying specifications of the ECS to be added. Create an AS configuration, for example, **as-config-discuz**.
		Creating an AS group	An AS group is the basis for performing scaling actions. Create an AS group, for example, **as-group-discuz**.

Step	Category	Substep	Description
		Creating an AS policy	An AS policy triggers scaling actions. Configure two CPU usage alarm policies for increasing or reducing the number of ECSs when website traffic rises or drops.
		Manually adding an ECS to an AS group	To ensure that discuz02 and the ECSs to be added to the AS group carry forum services together, manually add discuz02 to the AS group.
		Configuring **Min. Instances**	**Min. Instances** defines the minimum number of ECSs in an AS group. When the parameter is set to **1**, there is at least one ECS in the AS group. The ECS discuz02 is manually added and has the lowest priority to be removed when **Instance Removal Policy** is configured. Therefore, configuring **Min. Instances** prevents discuz02 from being removed.
Visiting the forum website	Verifying the configuration	Checking whether the forum website can be accessed	Obtain the EIP bound to the load balancer and visit http://*EIP*/forum.php. If the forum website is accessible, the configurations have taken effect.

Preparations

Follow the instructions provided in **Table 1-1** to set up the website. For details, see **Setting Up a Discuz! Forum**.

Creating an AS Configuration

An AS configuration specifies the specifications of ECS instances to be added. To enable the instances to automatically carry web services, use the image discuz_centos6.5 (40 GB) and ensure the parameter settings in the AS configuration the same as those of discuz02.

1. Log in to the management console. Under **Compute**, click **Auto Scaling**.
2. On the **Instance Scaling** page, click **Create AS Configuration**.

   Configure parameters listed in **Table 1-2**. Retain default settings for other parameters.

Table 1-2 Parameters required for creating an AS configuration

Parameter	Description	Example Value
Configuration Template	Select **Create new template** and configure the parameters, such as ECS type, vCPUs, memory, image, and disk, to create an AS configuration.	Create new template
Specifications	You can select multiple flavors to minimize the probability of capacity expansion failures due to insufficient resources of a flavor. Set **Flavor selection policy** to **Sequenced** or **Cost-centric** as required.	s3.medium.2 s3.large.2
Image	Specifies the software and system configuration template for the instances in an AS group.	discuz_centos6.5 (40 GB)
Disk	Stores data and manages the stored data for the instances in an AS group.	System disk: high I/O, 40 GB Data disk: high I/O, 100 GB
Security Group	Controls ECS access within or between security groups by defining access rules. Select security group **sg-DISCUZ**.	sg-DISCUZ
EIP	Not required if you have configured a load balancer for an AS group. The system automatically associates instances in the AS group to the load balancer. These instances will provide services via the EIP bound to the load balancer.	Do not use

3. After setting the parameters, click **Create Now**.

Creating an AS Group

1. Click **Create AS Group**.

   Configure parameters listed in **Table 1-3**. Retain default settings for other parameters.

Table 1-3 Parameters required for creating an AS group

Parameter	Description	Example Value
Max. Instances	Specifies the maximum number of instances in an AS group.	50
Expected Instances	Specifies the expected number of instances in an AS group. The ECSs in this practice are manually added to the AS group. To prevent scaling actions before the manually adding, set **Expected Instances** to **0**.	0
Min. Instances	Specifies the minimum number of instances in an AS group.	0
VPC	Provides the network used by instances in an AS group. Ensure that the parameter value is the same as the VPC in which discuz02 is deployed.	VPC-DISCUZ
Subnet	Manages networks in the VPC. Select the subnet created when you apply for the VPC.	vpc-test
Load Balancing	Evenly distributes traffic to instances in an AS group. A backend port is a service port on which a backend ECS listens for traffic, for example, set **Backend Port** to **80** and **Weight** to **1**.	Enhanced load balancer
Health Check Method	**ELB health check** is recommended, in which heartbeat messages are sent to backend ECSs for check.	ELB health check

2. After setting the parameters, click **Create Now**.
3. Back to the AS group list. The AS group is successfully created if its status changes to **Enabled**.

Creating an AS Policy

To automatically scale ECSs, configure two alarm policies to monitor CPU usage. One (**as-policy-discuz01**) is used to increase the number of ECSs when the website traffic rises, and the other (**as-policy-discuz02**) is used to reduce the number of ECSs when the website traffic drops.

1. Locate the row containing the created AS group **as-group-discuz** and click **View AS Policy** in the **Operation** column.

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

   Configure parameters listed in **Table 1-4** for **as-policy-discuz01**. When the system detects that the CPU usage exceeds 70% for three consecutive times, **as-policy-discuz01** is triggered and an ECS is added to the AS group.

   Table 1-4 Key parameters for creating AS policy **as-policy-discuz01**

Parameter	Description	Example Value
Policy Name	Specifies the name of the AS policy to be created.	as-policy-discuz01
Policy Type	Select **Alarm**.	Alarm
Alarm Rule	Specifies whether a new alarm rule is to be created (**Create**) or an existing alarm rule will be used (**Use existing**).	Create
Rule Name	Specifies the name of the alarm rule.	as-alarm-cpu-01
Monitoring Type	Specifies the type of monitoring metrics, which can be **System monitoring** or **Custom monitoring**. Select **System monitoring**.	System monitoring
Trigger Condition	Select monitoring metrics supported by AS and set alarm conditions for the metrics.	CPU Usage Max. >70%
Monitoring Interval	Specifies the interval at which the alarm status is updated based on the alarm rule.	5 minutes
Consecutive Occurrences	Specifies the number of sampling points when an alarm is triggered.	3

Parameter	Description	Example Value
Scaling Action	Specifies an action and the number or percentage of instances. The following scaling action options are available: • Add Adds instances to an AS group when the scaling action is performed. • Reduce Removes instances from an AS group when the scaling action is performed. • Set to Sets the expected number of instances in an AS group to a specified value.	Add 1 instance
Cooldown Period	To prevent an alarm-based policy from being triggered repeatedly by the same event, configure a cooldown period.	900

3. Click **OK**.
4. Click **Add AS Policy** again and create AS policy **as-policy-discuz02**. When the system detects that the CPU usage is lower than 30% for three consecutive times, **as-policy-discuz02** is triggered and an ECS is removed from the AS group.
5. Click **OK**.
6. Back to the AS policy list. The AS policies are successfully created if their statuses change to **Enabled**.

Manually Adding an ECS to an AS Group

Perform the following operations to manually add ECS **discuz02** to the AS group:

1. Click the name of the AS group **as-group-discuz** to switch to the page providing details about the AS group.
2. Click the **Instances** tab and manually add **discuz02** to the AS group.

Changing the Minimum Number of Instances

To ensure that **discuz02** is not removed from the AS group, perform the following operations to change the minimum number of instances:

1. Click the name of the AS group **as-group-discuz** to switch to the page providing details about the AS group.

2. Click **Modify AS Group** in the upper right corner of the page. Set **Min. Instances** to **1**.
3. Click **OK**.

Verifying Configurations

Check whether the forum website can be used. If the CPU usage of ECSs in the AS group remains higher than 70%, as shown on the **Monitoring** tab of the page providing details about the AS group, an ECS will be automatically added to the AS group (shown on the **Scaling Actions** tab). If the CPU usage remains lower than 30% and the AS group contains at least two ECSs, an ECS will be automatically removed from the AS group. If not, contact technical support to locate the fault.

Practice Extensions

- To deploy new applications on ECSs, use AS lifecycle hooks to perform customized operations on the instances to be added to or removed from an AS group. For details, see **Managing Lifecycle Hooks**.
- To modify the specifications of ECSs in an AS group, create a new AS configuration first. For details, see **Creating an AS Configuration from a New Specifications Template**. Then, replace the AS configuration used by the AS group with the one you created. For details, see **Changing the AS Configuration for an AS Group**.

2 Automatically Binding EIPs During Capacity Expansion and Releasing EIPs During Reduction

Background

- If no EIPs are assigned to the ECSs added to an AS group, the ECSs cannot provide public access. The system can automatically assign EIPs to the new ECSs during capacity expansion to keep them accessible.
- If you want to release ECS resources (for example, during off-peak hours) or no longer need public network access, the system can automatically release the EIPs (if any) during capacity reduction to optimize resource usage and reduce costs.

NOTE

An EIP is a public IP address that can be purchased and held independently. It can be bound to or unbound from an ECS. For details about EIP, see **What Are EIPs?**

Preparations

You have created an AS group and enabled it. For details, see **Creating an AS Group**.

Automatically Binding EIPs to ECSs During Capacity Expansion

1. Modify the AS configuration of the AS group.
   a. Log in to the Huawei Cloud console.
   b. On the console, choose **Compute** > **Auto Scaling** > **Instance Scaling**.
   c. On the **Instance Scaling** page, locate the AS group and choose **More** > **Modify AS Configuration** in the **Operation** column. Configure parameters listed in **Table 2-1**.

      NOTE

      This section uses the following parameters as an example. Retain the default values for the parameters that are not listed. For details, see **Using a New Specifications Template to Create an AS Configuration**.

Table 2-1 Parameters for configuring the AS configuration

Parameter	Example Value	Description
Name	as-config-test	Enter a name for the AS configuration as instructed.
Billing Mode	Pay-per-use	Pay-per-use is a post-paid charging mode, so the total expenditure depends directly on ECS usage frequency and duration. You can create or delete your ECSs at any time.
Flavor	s3.small.1	You can select a maximum of 10 flavors. You are advised to select several similar flavors based on the vCPU, memory, bandwidth, or the maximum number of packets that an ECS can transmit and receive per second.
Image	test	The image is used to deploy ECSs. You are advised to select the custom image where you have deployed your applications.
System Disk	General Purpose SSD, 100 GiB	Select a system disk for ECSs based on service requirements.
Security Group	sg-bp18kz60mefsicfg****	Select the security group you have created. For details about how to create a security group, see **Creating a Security Group**.
EIP	Automatically assign	The system automatically assigns an EIP with a dedicated bandwidth to each ECS. The bandwidth is configurable.

   d. Click **OK**.

2. Set a capacity expansion policy to trigger a capacity expansion action. For details, see **Creating an AS Policy**.
3. Check whether the new ECSs have EIPs bound.

   In the instance list of the AS group, click the ID of each created ECS. On the ECS details page, check whether the ECS has an EIP bound.

Releasing EIPs from ECSs During Capacity Reduction

1. Modify the AS group.
   a. Log in to the Huawei Cloud management console.
   b. On the console, choose **Compute** > **Auto Scaling** > **Instance Scaling**.
   c. On the **Instance Scaling** page, locate the AS group and choose **More** > **Modify AS Group** in the **Operation** column. Configure parameters listed in **Table 2-2**.

      NOTE

      This section uses the following parameters as an example. Retain the default values for the parameters that are not listed. For details, see **Creating an AS Group**.

      Table 2-2 Parameters for configuring the AS group

Parameter	Example Value	Description
EIP	Release	If you select **Release**, the system releases the EIP from the ECS when an ECS is removed from the AS group. Otherwise, the system reserves the EIP.

   d. Click **OK**.
2. Set a capacity reduction policy to trigger a capacity reduction action. For details, see **Creating an AS Policy**.
3. Check whether the EIPs of the ECSs removed from the AS group have been released.

   You can go to the EIP list to check if they are displayed. If the EIPs do not exist, they have been released.

3 Configuring Scaling Policies and Multiple Flavors to Reduce Costs

- When creating an AS group, you can configure a multi-AZ scaling policy. If resources in an AZ are insufficient or the AZ is faulty, Auto Scaling can schedule resources from other AZs to ensure high service availability.
- When creating an AS configuration, you can select multiple flavors to prevent scaling action failures due to insufficient inventory of a single flavor.

  NOTE

  - **Scaling Policy**

    You can specify multiple flavors when creating an AS configuration. If the inventory of an individual flavor in an AZ is insufficient, Auto Scaling automatically attempts to create instances in other AZs to ensure that scaling actions can be executed. You can configure scaling policies (such as the cost-centered policy) based on service deployment to meet service requirements. If a scaling group uses the cost-centered policy, Auto Scaling attempts to create ECSs in ascending order of the flavor unit price.

  - **Multiple Flavors**

    You can select multiple flavors and set **Flavor selection policy** to **Cost-centered** in a scaling configuration. During AS group expansion, the flavor with the minimum cost comes first. When creating an ECS in an AS group, the system selects the flavor with the minimum cost. If the ECS cannot be created, the system selects one with the minimum cost from the remained flavors, and so on. Multiple flavors can effectively handle the insufficient inventory issue of a single flavor and ensure that scaling actions can be executed. For example, during peak hours, you may need to quickly obtain ECSs with different flavors to carry service traffic and deliver high performance, rather than focusing on one specific flavor. In this case, multiple flavors can flexibly meet the requirements.

This section describes how to configure multi-AZ scaling, multiple flavors, and cost-centered policies to effectively improve the scaling success rate and reduce costs.

Prerequisites

You have created a VPC with a subnet. For details, see **Creating a VPC with a Subnet**.

Procedure

1. Create an AS configuration.
   a. Log in to the Huawei Cloud management console.
   b. On the console, choose **Compute** > **Auto Scaling** > **Instance Scaling**.
   c. Click **Create AS Configuration** and configure parameters as instructed.

      NOTE

      This section uses the following parameters as an example. Retain the default values for the parameters that are not listed. For details, see **Using a New Specifications Template to Create an AS Configuration**.

      Table 3-1 Parameters for configuring the AS configuration

Parameter	Example Value	Description
Name	as-config-test	Enter a name for the AS configuration.
Billing Mode	Pay-per-use	You are billed for how long you use each ECS. You can create or delete such an ECS at any time.
Flavor	s3.small.1, s3.medium.2, s3.large.2 (The price per unit of each flavor increases in ascending order.)	A maximum of 10 flavors can be selected. You are advised to select several similar flavors based on the vCPU, memory, bandwidth, or the maximum number of packets that an ECS can transmit and receive per second.
Flavor selection policy	Cost-centered	When creating an instance in an AS group, the system selects the flavor with the minimum cost. If the instance cannot be created, the system selects one with the minimum cost from the remained flavors, and so on.
Image	test	Select an image used to deploy ECSs. You are advised to select the custom image where you have deployed your applications.
System Disk	General Purpose SSD, 100 GiB	Select a system disk for ECSs based on service requirements.

Parameter	Example Value	Description
Security Group	sg-bp18kz60mefsicfg****	Select the security group you have created. For details about how to create a security group, see **Creating a Security Group**.

d. Click **Create Now**.

2. Create an AS group.

a. On the **Instance Scaling** page, click **Create AS Group**.

b. Configure parameters as instructed.

NOTE

This section uses the following parameters as an example. Retain the default values for the parameters that are not listed. For details, see **Creating an AS Group**.

Table 3-2 Parameters for configuring the AS group

Parameter	Example Value	Description
Name	test	Enter the AS group name as instructed.
AZ	AZ1, AZ2, AZ3	Select the AZs where ECSs will be deployed.
Multi-AZ Scaling Policy	Balanced	**Balanced**: When an ECS is added for capacity expansion, the target AZ is determined for the balanced number of ECSs in each AZ. If the attempt fails, the target AZ is determined in the order in which they were selected.
Min. Instances	0	Set the minimum number of instances in an AS group. If the number of instances in an AS group is less than the minimum number, the AS group will automatically add instances until the number of instances reaches the minimum number.

Parameter	Example Value	Description
Max. Instances	120	Set the maximum number of instances in an AS group. If the number of instances in an AS group exceeds the maximum number, the AS group will automatically remove instances until the number of instances reaches the maximum number.
Expected Instances	0	Specifies the expected number of ECSs in the AS group. After the creation, you can change this value. However, changing this value will trigger a scaling action. The value cannot be less than the number of minimum instances or greater than the number of maximum instances.
AS Configuration	as-config-test	Select the AS configuration you created for the AS group.
VPC	Vpc-test	All ECSs in the AS group are deployed in this VPC.
Subnet	Subnet-test	A maximum of five subnets can be selected. The AS group automatically attach all NICs to the created ECSs. By default, the first subnet you select is used as the primary NIC of the ECS, and other subnets are used as the extension NICs.
Health Check Method	ECS health check	Checks the ECS running status. If an ECS is stopped or deleted, it is considered to be unhealthy.
Health Check Interval	5 minutes	Specifies the health check interval of the AS group. You can set a health check interval, such as 10 seconds, 1 minute, 5 minutes, 15 minutes, 1 hour, or 3 hours, based on service requirements.

c. After setting the parameters, click **Create Now**.

d. Return to the AS group list. The AS group is successfully created if its status is **Enabled**.

3. Configure an AS policy to trigger a capacity expansion action. For details, see **Creating an AS Policy**.

Verification

- **Expected result**: When the scaling action is triggered, the AS group preferentially creates an ECS using s3.small.1. If the ECS cannot be created in AZ 1, AZ 2, and AZ 3, the AS group attempts the flavor s3.medium.2 and then s3.large.2 in sequence.
- **Verification**: After the capacity expansion action is triggered, an ECS is added to the AS group. On the **Instances** tab, click the name of the added ECS. On the ECS details page, view the ECS flavor. If the ECS flavor is s3.small.1, the AS policy is in effect. The multi-AZ scaling and cost-centered policies configured for the AS group can significantly reduce costs.

4 Deploying an HA Compute Cluster Using the Balanced Policy

Scenarios

An HA compute cluster is needed in scenarios such as distributed big data compute, AI training, or traffic surges. Auto Scaling can automatically batch create ECSs and evenly distribute them across multiple AZs using the **Balanced** policy. Auto Scaling also monitors the status of ECSs in real time to ensure the high availability (HA) of the cluster.

Advantages

- Zero O&M cost

  With Auto Scaling, ECSs can be automatically scaled in or out without manual intervention.

- HA

  The **Balanced** policy can automatically distribute ECSs across AZs. This prevents capacity expansion failures caused by insufficient inventory in a single AZ. Health check is enabled by default to ensure that all ECSs in an AS group are running properly.

This section describes how to use Auto Scaling to evenly distribute ECSs across AZs and deploy an HA compute cluster.

Step 1: Create an AS Configuration

NOTICE

- Evaluate service modules based on your service architecture and create an AS group for the service module that requires an HA compute cluster.
- Prepare a custom image with service applications deployed. When configuring the desired AS configuration, select the custom image to ensure that the automatically created ECSs meet the requirements. For details, see **Creating a Full-ECS Image from an ECS**.

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 Configuration** and configure parameters as instructed.

   NOTE

   This section uses the following parameters as an example. Retain the default values for the parameters that are not listed. For details, see **Using a New Specifications Template to Create an AS Configuration**.

   Table 4-1 Parameters for configuring the AS configuration

Parameter	Example Value	Description
Name	as-config-test	Enter a name for the AS configuration as instructed.
Billing Mode	Pay-per-use	Pay-per-use is a post-paid charging mode, so the total expenditures depend directly on ECS usage frequency and duration. You can create or delete your ECSs at any time.
Flavor	s3.small.1	Set the vCPU and memory of the ECSs to be created.
Image	test	Select an image used to deploy ECSs. You are advised to select the custom image where you have deployed your applications.
System Disk	General Purpose SSD, 100 GiB	Select a system disk for ECSs based on service requirements.
Security Group	sg-bp18kz60mefsicfg****	Select the security group you have created. For details about how to create a security group, see **Creating a Security Group**.

4. Click **Create Now**.

Step 2: Create an AS Group

1. On the **Instance Scaling** page, click **Create AS Group**.
2. Configure parameters as instructed.

   NOTE

   This section uses the following parameters as an example. Retain the default values for the parameters that are not listed. For details, see **Creating an AS Group**.

Table 4-2 Parameters for configuring the AS group

Parameter	Example Value	Description
Name	test	Enter the AS group name as instructed.
AZ	AZ1, AZ2, AZ3	Select the AZs where ECSs will be deployed.
Multi-AZ Scaling Policy	Balanced	This parameter is mandatory when two or more AZs are selected.
Min. Instances	10	Set the minimum number of instances in an AS group. If the number of instances in an AS group is less than the minimum number, the AS group will automatically add instances until the number of instances reaches the minimum number.
Max. Instances	120	Set the maximum number of instances in an AS group. If the number of instances in an AS group exceeds the maximum number, the AS group will automatically remove instances until the number of instances reaches the maximum number.
Expected Instances	10	Specifies the expected number of ECSs in an AS group. After the creation, you can change this value. However, changing this value will trigger a scaling action. The value cannot be less than the number of minimum instances or greater than the number of maximum instances.
AS Configuration	as-config-test	Select the AS configuration you created for the AS group.
VPC	Vpc-test	All ECSs in the AS group are deployed in this VPC.

Parameter	Example Value	Description
Subnet	Subnet-test	A maximum of five subnets can be selected. The AS group automatically attach all NICs to the created ECSs. By default, the first subnet you select is used as the primary NIC of the ECS, and other subnets are used as the extension NICs.
Health Check Method	ECS health check	Checks the ECS status. If an ECS is stopped or deleted, it is considered to be unhealthy.
Health Check Interval	5 minutes	Specifies the health check interval of the AS group. You can set a health check interval, such as 10 seconds, 1 minute, 5 minutes, 15 minutes, 1 hour, or 3 hours, based on service requirements.

3. After setting the parameters, click **Create Now**.
4. Back to the AS group list. The AS group is successfully created if its status is **Enabled**.

Result

The minimum number of instances in the AS group is set to 10. After the AS group is enabled, the AS group automatically deploys 10 ECSs evenly in the selected AZs to form a compute cluster. The AS group ensures the cluster HA and reduces resource costs in the following ways:

- If the inventory is insufficient in an AZ, the system automatically attempts to deploy ECSs in other AZs to reduce impact on services.
- The AS group automatically removes unhealthy ECSs and creates new ECSs.

5 Using Tag-based Authentication for Fine-grained Management of Auto Scaling

Background

You can achieve fine-grained management of Auto Scaling using tags together with Identity and Access Management (IAM).

- Tags are identifiers of cloud resources. They help you classify, search for, and aggregate cloud resources with the same feature and from different dimensions.
- IAM provides permission management. You can manage user identities and control access to and operations on cloud resources based on permission policies. For details, see **What Is IAM?**

This section describes how to use tag-based authentication to grant different permissions to different IAM users for improved management efficiency and reduced information leakage risks.

Example Scenarios

Assume that two AS groups have been created. The table below lists the AS groups and their tags.

Table 5-1 AS groups and their tags

AS Group	Name	Tag
AS group 1	asg-001	**team:game1**, where **team** is the tag key and **game1** is the tag value.
AS group 2	asg-002	**team:game2**, where **team** is the tag key and **game2** is the tag value.

You can use tag-based authentication to control the specific permissions of an IAM user on the preceding AS groups. The following are examples:

- **Scenario 1**: AS group 1 cannot be created unless **team:game1** is attached to it.
- **Scenario 2**: Only AS group 1 with tag **team:game1** attached can be queried.
- **Scenario 3**: AS group 1 and AS group 2 have tags **team:game1** and **team:game2** attached, respectively. Only the AS group 1 with tag **team:game1** attached can be operated.

Procedure

NOTICE

Ensure that you have created an IAM user and granted read-only permissions for ECS, VPC, and IMS to the IAM user. If no IAM user is available, create one. For details, see **Creating an IAM User**.

1. Create two AS groups. For details, see **Creating an AS Group**.
2. Log in to the IAM console.
3. Create a custom policy. For details, see **Creating a Custom Policy**.

   You can set multiple tag-based authentication conditions for cloud resources to restrict the operation permissions on Auto Scaling resources. The table below lists the supported tag-based authentication conditions.

   Table 5-2 Tag-based authentication conditions

Condition Key	Description
g:RequestTag	A specific tag must be transferred in a request. Otherwise, the authentication fails. If an API request does not contain a tag parameter, g:RequestTag cannot be used. Otherwise, the authentication fails.
g:ResourceTag	The specified resource must contain a specific tag. Otherwise, the authentication fails. If an API request does not contain a resource ID parameter, g:ResourceTag cannot be used. Otherwise, the authentication fails.

   - **Scenario 1: AS group1 cannot be created without a specific tag attached.**

     AS group 1 cannot be created unless **team:game1** is attached to it.

     The custom policy is as follows:

```
{
 "Version": "5.0",
 "Statement": [
  {
   "Effect": "Allow",
   "Action": [
```



```
      "as:scalingGroup:create"
    ],
    "Condition": {
      "StringEquals": {
        "g:RequestTag/team": [
          "game1"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "as:scalingConfig:list"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "as::listQuotas"
    ]
  }
 ]
}
```

- **Scenario 2: Only AS group 1 with a specific tag attached can be queried.**

  After AS group 1 has tag **team:game1** attached, only AS group 1 resources can be queried.

  The custom policy is as follows:

```
{
 "Version": "5.0",
 "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "as:scalingGroup:get"
    ],
    "Condition": {
      "StringEquals": {
        "g:ResourceTag/team": [
          "game1"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "as:scalingConfig:list"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "as::listQuotas"
    ]
  }
 ]
}
```

- **Scenario 3: Only AS group 1 can be operated.**

  AS group 1 and AS group 2 have tags **team:game1** and **team:game2** attached, respectively. Only the AS group 1 with tag **team:game1** attached can be operated.

  The custom policy is as follows:


```
{
  "Version": "5.0",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "as:scalingConfig:list"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "as::listQuotas"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "as:scalingGroup:update"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "as:scalingGroup:resume",
        "as:scalingGroup:pause"
      ],
      "Condition": {
        "StringEquals": {
          "g:ResourceTag/team": [
            "game1"
          ]
        }
      }
    }
  ]
}
```

4. Attach the custom policy to the IAM user you want to control access for. For details, see **Assigning Permissions to an IAM User**.

Verification

Verify whether the permission policy is applied.

- **Scenario 1: AS group1 cannot be created without a specific tag attached.**

  You have attached tag **team:game1** to AS group 1 and created AS group1 successfully. If tags other than **team:game1** is attached to AS group 1, the system displays a message indicating that you do not have required permissions to create AS group 1.

- **Scenario 2: Only AS group 1 with a specific tag attached can be queried.**

  If you query AS group 1 (with tag **team:game1** attached) without filtering tags, you can obtain the information about this AS group.

  If you query an AS group (without tag **team:game1** attached) other than AS group 1, no information is displayed.

- **Scenario 3: Only AS group 1 can be operated.**

  You have permissions to modify AS group 1 (with tag **team:game1** attached).

  If you attempt to modify AS group 2 with any other tag other than **team:game1** attached, the system displays a message indicating that you do not have required permissions to modify this AS group.

