

# Auto Scaling

## Descripción general del servicio

Edición 04  
Fecha 2021-10-30



**Copyright © Huawei Technologies Co., Ltd. 2023. Todos los derechos reservados.**

Quedan terminantemente prohibidas la reproducción y la divulgación del presente documento en todo o en parte, de cualquier forma y por cualquier medio, sin la autorización previa de Huawei Technologies Co., Ltd. otorgada por escrito.

## **Marcas y permisos**



HUAWEI y otras marcas registradas de Huawei pertenecen a Huawei Technologies Co., Ltd.

Todas las demás marcas registradas y los otros nombres comerciales mencionados en este documento son propiedad de sus respectivos titulares.

## **Aviso**

Las funciones, los productos y los servicios adquiridos están estipulados en el contrato celebrado entre Huawei y el cliente. Es posible que la totalidad o parte de los productos, las funciones y los servicios descritos en el presente documento no se encuentren dentro del alcance de compra o de uso. A menos que el contrato especifique lo contrario, ninguna de las afirmaciones, informaciones ni recomendaciones contenidas en este documento constituye garantía alguna, ni expresa ni implícita.

La información contenida en este documento se encuentra sujeta a cambios sin previo aviso. En la preparación de este documento se realizaron todos los esfuerzos para garantizar la precisión de sus contenidos. Sin embargo, ninguna declaración, información ni recomendación contenida en el presente constituye garantía alguna, ni expresa ni implícita.

---

# Índice

---

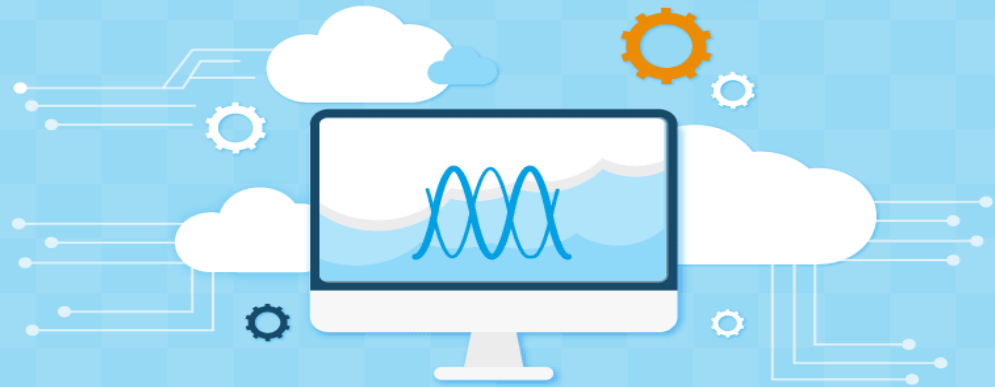
<b>1 AS Infografías.....</b>	<b>1</b>
<b>2 ¿Qué es Auto Scaling?.....</b>	<b>3</b>
<b>3 Ventajas de AS.....</b>	<b>5</b>
<b>4 Ciclo de vida de instancia.....</b>	<b>10</b>
<b>5 Restricciones.....</b>	<b>15</b>
<b>6 Región y AZ.....</b>	<b>17</b>
<b>7 Facturación.....</b>	<b>19</b>
<b>8 AS y otros servicios.....</b>	<b>20</b>
<b>9 Gestión de permisos.....</b>	<b>23</b>
<b>10 Conceptos básicos.....</b>	<b>26</b>
<b>11 Historial de revisiones.....</b>	<b>28</b>

# 1 AS Infografías

---



# Auto Scaling

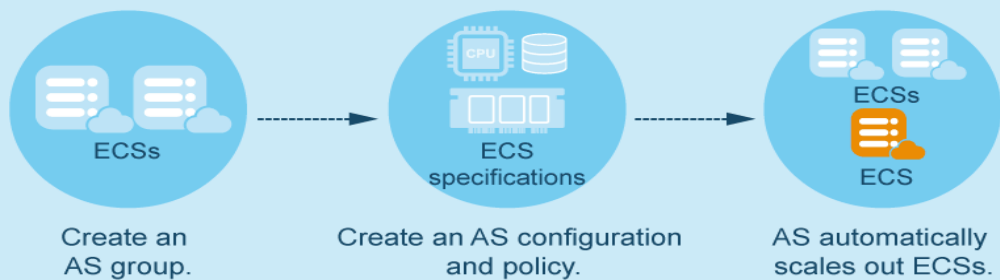


## 01

### What Is AS?

Auto Scaling (AS) automatically adjusts resources to keep up with changes in demand based on pre-configured AS policies.

You can specify AS configurations and policies based on service requirements. These configurations and policies free you from having to repeatedly adjust resources to keep up with service changes and spikes in demand. In this way, AS helps you reduce the resources and manpower required.



**?** During the 11.11 Shopping Festival in China on November 11, it is extremely hard to keep up with the massive increase in demand. Traditional solutions are not up to the task.

## 02 Advantages of AS

Provision as many as servers as required for peak demand. All that capacity will be wasted during off-peak hours.

Provision servers based on average loads of applications. Capacity will be insufficient during peak hours.

AS perfectly resolves this issue.

# 2 ¿Qué es Auto Scaling?

---

## Introducción al AS

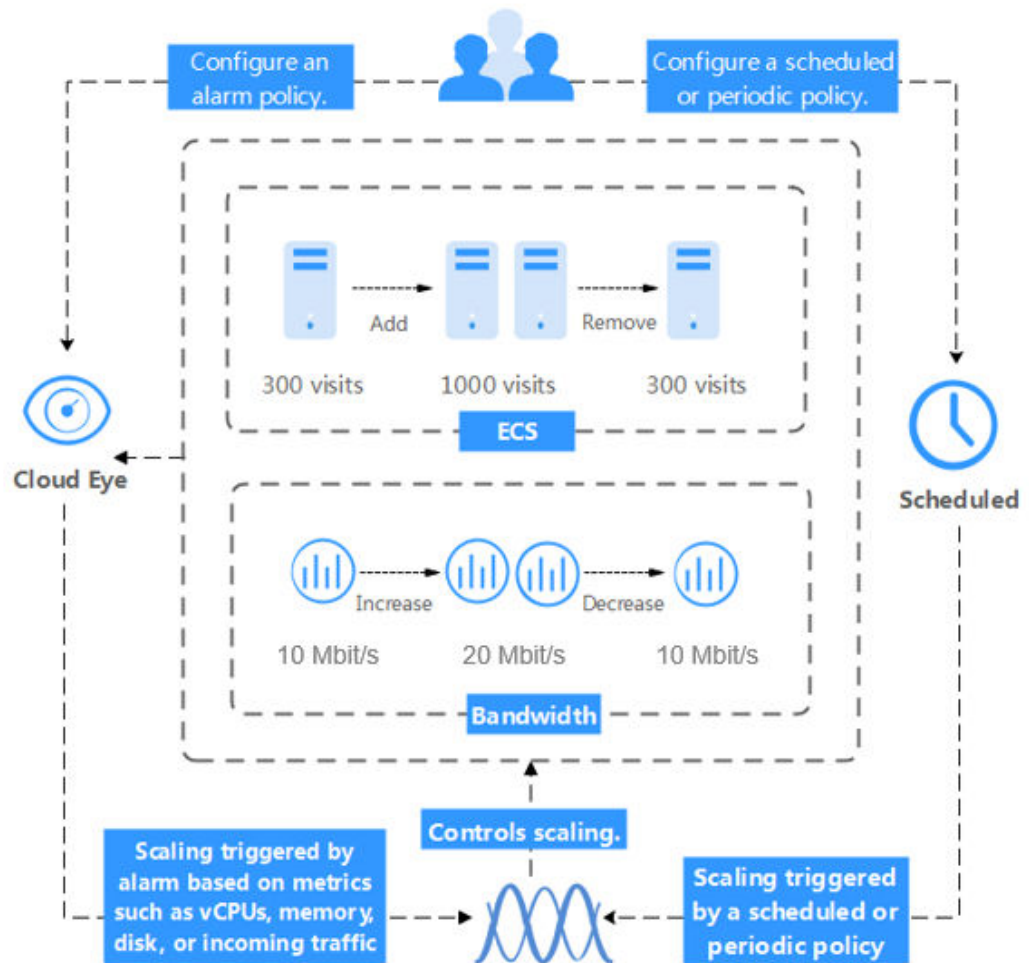
Auto Scaling (AS) le ayuda a escalar automáticamente Elastic Cloud Server (ECS) y los recursos de ancho de banda para mantenerse al día con los cambios en la demanda basados en políticas de AS preconfiguradas. Le permite agregar instancias de ECS o aumentar los anchos de banda para manejar los aumentos en la carga y también ahorrar dinero al eliminar recursos que están inactivos.

## Arquitectura

AS le permite escalar instancias de ECS y anchos de banda.

- **Control de ajuste:** puede configurar políticas de AS, configurar umbrales de métrica y programar cuando se realizan diferentes acciones de ajuste. AS activará acciones de ajuste en una programación repetitiva, en un momento específico o cuando se alcancen los umbrales configurados.
- **Configuración de políticas:** puede configurar políticas basadas en alarmas, programadas y periódicas según sea necesario.
- **Políticas basadas en alarmas:** puede configurar las acciones de ajuste para que se realicen cuando las métricas de alarma como vCPU, memoria, disco y tráfico entrante alcancen los umbrales.
- **Políticas programadas:** puede programar acciones de ajuste para que se realicen en un momento específico.
- **Políticas periódicas:** puede configurar las acciones de ajuste para que se realicen a intervalos programados, una hora específica o dentro de un intervalo de tiempo determinado.
- Cuando Cloud Eye genera una alarma para una métrica de monitorización, por ejemplo, el uso de la CPU, AS aumenta o disminuye automáticamente el número de instancias en el grupo AS o los anchos de banda.
- Cuando llega el tiempo de activación configurado, se activa una acción de ajuste para aumentar o disminuir el número de instancias de ECS o los anchos de banda.

Figura 2-1 Arquitectura de AS



## Acceso a AS

La nube pública proporciona una plataforma de gestión de servicios basada en web. Puede acceder a AS mediante interfaces de programación de aplicaciones (API) compatibles con HTTPS o la consola de gestión.

- **Llamar a API**  
Utilice este método si necesita integrar AS en la plataforma de nube pública en un sistema de terceros para el desarrollo secundario. Para obtener más información, consulte [Referencia de la API de Auto Scaling](#).
- **Consola de gestión**  
Utilice este método si no necesita integrar AS con un sistema de terceros. Después de registrarse en la nube pública, inicie sesión en la consola de gestión y seleccione **Auto Scaling** en la lista de servicios en la página de inicio.

# 3 Ventajas de AS

---

AS escala automáticamente los recursos para mantenerse al día con las demandas de servicio basadas en políticas de AS preconfiguradas. Con el ajuste automático de recursos, puede disfrutar de costos reducidos, disponibilidad mejorada y alta tolerancia a fallos. AS se utiliza para los siguientes escenarios:

- Foros con mucho tráfico: El tráfico en un foro popular es difícil de predecir. AS ajusta dinámicamente el número de instancias de ECS basándose en métricas de ECS monitorizadas, como vCPU y uso de memoria.
- Comercio electrónico: Durante las grandes promociones, los sitios web de comercio electrónico necesitan más recursos. AS aumenta automáticamente las instancias de ECS y los anchos de banda en cuestión de minutos para garantizar que las promociones se desarrollen sin problemas.
- Transmisión en vivo: Un sitio web de transmisión en vivo puede transmitir programas populares de 14:00 a 16:00 todos los días. AS escala automáticamente los recursos de ECS y de ancho de banda durante este período para garantizar una experiencia fluida del espectador.

## Ajuste automático de recursos

AS agrega instancias de ECS y aumenta el ancho de banda para sus aplicaciones cuando el volumen de acceso aumenta y elimina los recursos innecesarios cuando el volumen de acceso disminuye, lo que garantiza la estabilidad y disponibilidad del sistema.

- Escalado de instancias de ECS bajo demanda

AS escala las instancias de ECS para aplicaciones según la demanda, lo que mejora la gestión de costos. Las instancias de ECS se pueden escalar dinámicamente, según una programación o manualmente:

  - Ajuste dinámico

El ajuste dinámico permite escalar recursos en respuesta a la demanda cambiante mediante políticas basadas en alarmas.
  - Ajuste programado

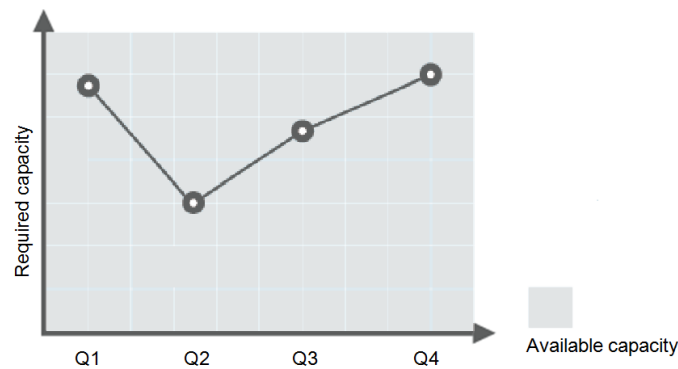
El ajuste programado le ayuda a configurar su propio programa de ajuste de acuerdo con los cambios de carga predecibles mediante la creación de políticas periódicas o programadas.
  - Ajuste manual



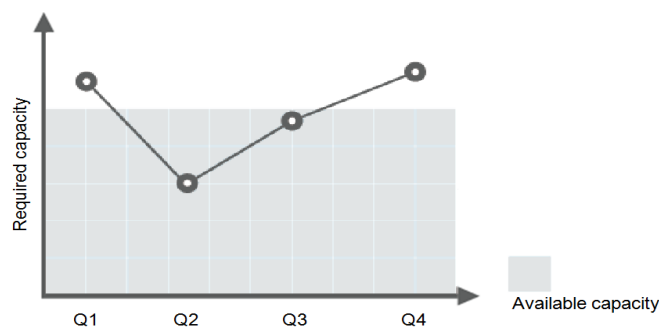
Puede cambiar manualmente el número esperado de instancias del grupo AS, o bien agregar o quitar instancias al grupo AS o del grupo AS.

Considere una aplicación de reserva de billetes de tren que se ejecuta en la nube pública. La carga de la aplicación puede ser relativamente baja durante Q2 y Q3 porque no hay muchos viajeros, pero relativamente alta durante Q1 y Q4. Tradicionalmente, hay dos maneras de planificar estos cambios en la carga. La primera opción es aprovisionar servidores suficientes para que la aplicación siempre tenga suficiente capacidad para satisfacer la demanda, como se muestra en **Figura 3-1**. La segunda opción es aprovisionar servidores de acuerdo con la carga media de la aplicación, como se muestra en **Figura 3-2**. Sin embargo, estas dos opciones pueden desperdiciar recursos o no poder satisfacer la demanda durante las temporadas altas. Al habilitar AS para esta aplicación, usted tiene una tercera opción disponible. AS le ayuda a escalar servidores para mantenerse al día con los cambios en la demanda. Esto permite que la aplicación mantenga un rendimiento estable y predecible sin desperdiciar dinero en recursos innecesarios, como se muestra en **Figura 3-3**.

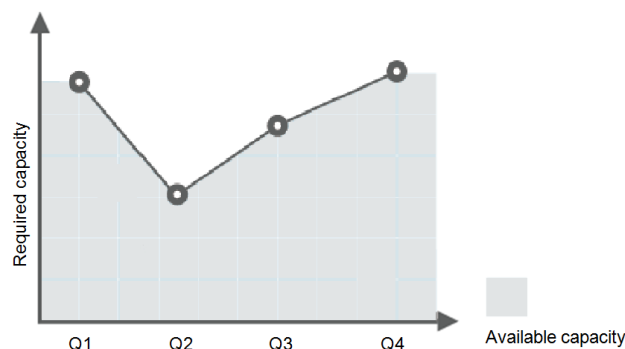
**Figura 3-1** Capacidad sobreaprovisionada



**Figura 3-2** Capacidad insuficiente



**Figura 3-3** Capacidad autoescalada



- Escalado del ancho de banda bajo demanda  
AS ajusta el ancho de banda para una aplicación según la demanda, reduciendo los costos de ancho de banda.  
Existen tres tipos de políticas de ajuste que puede utilizar para ajustar el ancho de banda IP a petición:
  - Políticas basadas en alarmas  
Puede configurar activadores basados en métricas como el tráfico saliente y el ancho de banda. Cuando el sistema detecta que se cumplen las condiciones de activación, el sistema ajusta automáticamente el ancho de banda.
  - Políticas programadas  
El sistema aumenta, disminuye o ajusta automáticamente el ancho de banda a un valor fijo en una programación fija.
  - Políticas periódicas  
El sistema ajusta periódicamente el ancho de banda basándose en un ciclo periódico configurado.

Por ejemplo, puede utilizar una política basada en alarmas para regular el ancho de banda de un sitio web de transmisión en vivo.

Para un sitio web de transmisión en vivo, la carga del servicio es difícil de predecir. En este ejemplo, el ancho de banda necesita ajustarse dinámicamente entre 10 Mbit/s y 30 Mbit/s basándose en métricas tales como tráfico saliente y tráfico entrante. AS puede ajustar automáticamente el ancho de banda para cumplir con los requisitos. Solo tiene que seleccionar el EIP relevante y crear dos políticas de alarma. Una política es aumentar el ancho de banda en 2 Mbit/s cuando el tráfico saliente es mayor que  $X$  bytes, con el límite establecido en 30 Mbit/s. La otra política es reducir el ancho de banda en 2 Mbit/s cuando el tráfico saliente es menor que  $X$  bytes, con el límite establecido en 10 Mbit/s.

- Distribuir instancias uniformemente por AZ  
Para reducir el impacto de la interrupción de la alimentación o de la red en la estabilidad del sistema, AS intenta distribuir las instancias ECS de manera uniforme entre las AZ que utiliza un grupo AS.

Una región es un área geográfica donde se encuentran los recursos utilizados por las instancias de ECS. Cada región contiene varias zonas de disponibilidad (AZ) donde los recursos utilizan fuentes de alimentación y redes independientes. Las AZ están físicamente aisladas entre sí pero interconectadas a través de una intranet. Las AZ están diseñadas para ser aisladas de fallos en otras AZ. Proporcionan conexiones de red rentables y de baja latencia a otras AZ de la misma región.

Un grupo AS puede contener instancias de ECS en una o más AZ dentro de una región. Durante la escala de la capacidad de un grupo AS, AS intenta distribuir uniformemente

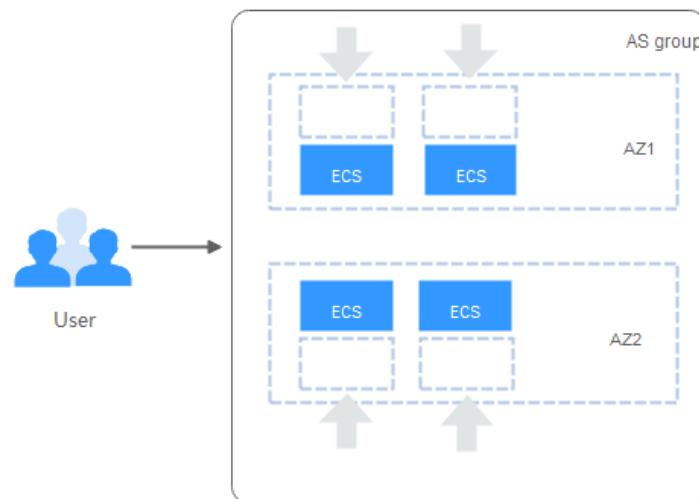
las instancias de ECS entre las AZ utilizadas por el grupo AS basándose en las siguientes reglas:

Distribuir uniformemente nuevas instancias a AZ balanceadas

AS intenta distribuir uniformemente las instancias de ECS a través de las AZ utilizadas por un grupo de AS. Para hacerlo, AS agrega nuevas instancias a la AZ con el menor número de instancias.

Considere un grupo AS que contiene cuatro instancias que están distribuidas uniformemente en las dos AZ utilizadas por el grupo AS. Si se activa una acción de ajuste para agregar cuatro instancias más al grupo AS, AS agrega dos a cada AZ.

**Figura 3-4** Distribución uniforme de instancias

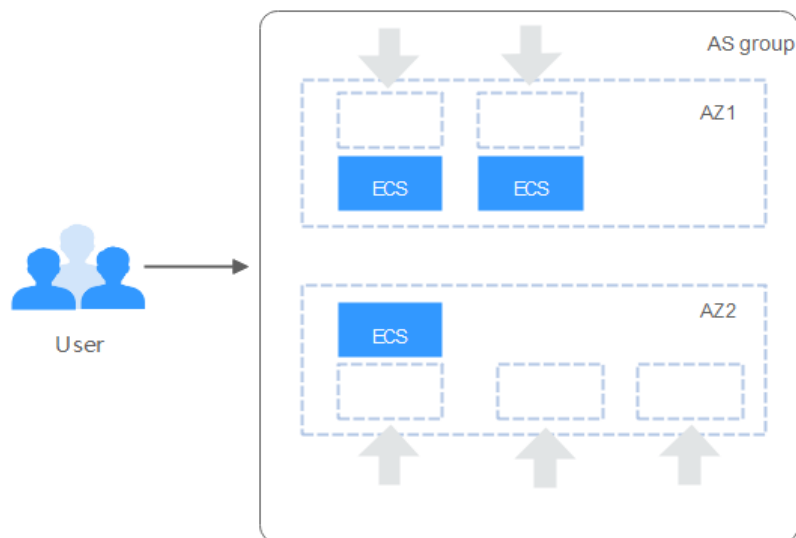


Reequilibrio de instancias entre las AZ

Después de agregar o quitar instancias manualmente a o desde un grupo AS, el grupo AS puede desequilibrarse entre las AZ. AS compensa reequilibrando las AZ durante la siguiente acción de ajuste.

Considérese un grupo AS que contiene tres instancias que están distribuidas en AZ 1 y AZ 2, con dos en AZ 1 y uno en AZ 2. Si se activa una acción de ajuste para agregar cinco instancias más al grupo AS, AS agrega dos a AZ 1 y tres a AZ 2.

**Figura 3-5** Reequilibrio de instancias



## Gestión de costes mejorada

AS le permite usar instancias de ECS y anchos de banda bajo demanda escalando automáticamente los recursos para sus aplicaciones, eliminando el desperdicio de recursos y reduciendo los costos.

## Mayor disponibilidad

AS garantiza que siempre tenga la cantidad adecuada de recursos disponibles para manejar la carga fluctuante de sus aplicaciones.

### Uso de ELB con AS

Al trabajar con ELB, AS escala automáticamente las instancias de ECS en función de los cambios en la demanda, al tiempo que garantiza que la carga de todas las instancias en un grupo de AS se mantenga equilibrada.

Después de que se habilita ELB para un grupo AS, AS asocia automáticamente un oyente de equilibrio de carga con cualquier instancia agregada al grupo AS. A continuación, ELB distribuye automáticamente el tráfico a todas las instancias sanas del grupo AS a través del oyente, lo que mejora la disponibilidad del sistema. Si las instancias del grupo AS ejecutan un rango de diferentes tipos de aplicaciones, puede vincular varios oyentes de equilibrio de carga al grupo AS para escuchar cada una de estas aplicaciones, mejorando la escalabilidad del servicio.

## Alta tolerancia a fallos

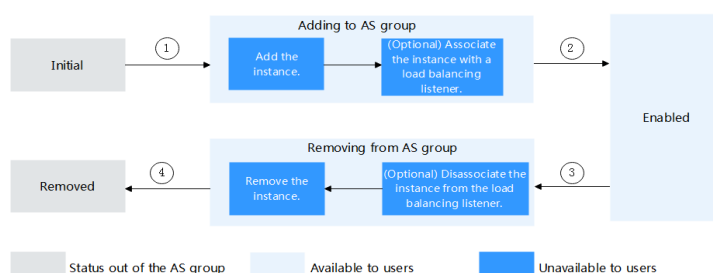
AS supervisa las instancias de un grupo AS y reemplaza las instancias no saludables que detecte por otras nuevas. Además, AS distribuye de manera uniforme las instancias entre las AZ para garantizar la estabilidad del sistema.

# 4 Ciclo de vida de instancia

Una instancia de ECS en un grupo AS pasa por diferentes estatus desde su creación hasta su eliminación.

El estado de la instancia cambia como se muestra en **Figura 4-1** si no ha agregado un enlace de ciclo de vida al grupo AS.

**Figura 4-1** Ciclo de vida de instancia



Cuando se cumple la condición de activación 2 o 4, el sistema pone las instancias de forma autónoma en el siguiente estado.

**Tabla 4-1** Estado de instancia

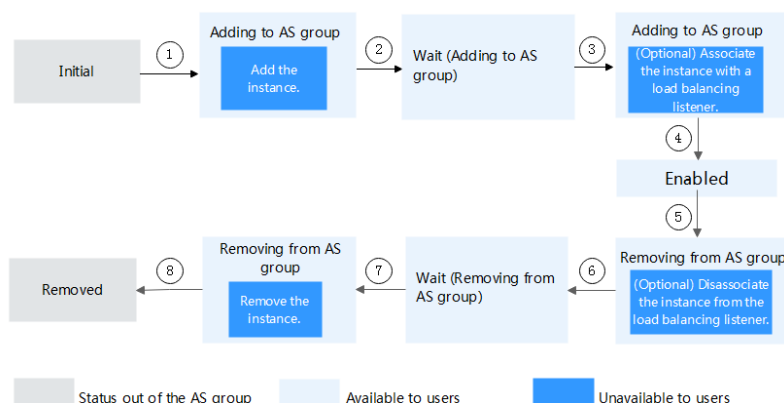
Estado	Subestatos	Descripción del estado	Condición de activación
Initial	No hay	La instancia no se ha agregado al grupo de AS.	El estado de la instancia se cambiará a <b>Adding to AS group</b> cuando se cumpla alguna de las siguientes condiciones: <ul style="list-style-type: none"> <li>● Aumenta manualmente el número esperado de instancias del grupo AS.</li> </ul>
Adding to AS group	Adición de la instancia.	Cuando se cumple la condición 1 de activación, AS agrega la instancia para ampliar la capacidad del grupo AS.	

Estado	Subestados	Descripción del estado	Condición de activación
	(Opcional) Asocie la instancia con un oyente de equilibrio de carga.	Cuando se cumple la condición 1 de activación, AS asocia la instancia creada con el oyente de equilibrado de carga.	<ul style="list-style-type: none"> <li>● El sistema amplía automáticamente la capacidad del grupo AS.</li> <li>● Puede agregar instancias manualmente al grupo AS.</li> </ul>
Enabled	No hay	La instancia se agrega al grupo AS y comienza a procesar el tráfico de servicio.	El estado de la instancia cambia de <b>Enabled</b> a <b>Removing from AS group</b> cuando se cumple cualquiera de las siguientes condiciones:
Removing from AS group	(Opcional) Disocie la instancia del oyente de equilibrio de carga.	Cuando se cumple la condición de activación 3, el grupo AS comienza a reducir recursos y disociar la instancia del oyente de equilibrado de carga.	<ul style="list-style-type: none"> <li>● Se reduce manualmente el número esperado de instancias del grupo AS.</li> <li>● El sistema elimina automáticamente instancias en una acción de ajuste.</li> <li>● Una comprobación de estado muestra que una instancia habilitada no está en estado y el sistema la elimina del grupo AS.</li> <li>● Eliminar instancias manualmente del grupo AS.</li> </ul>
	Retire la instancia.	Después de que las instancias se desvinculan del oyente de equilibrio de carga, se eliminan del grupo AS.	
Removed	No hay	El ciclo de vida de la instancia en el grupo AS finaliza.	No hay

Cuando se agrega una instancia de ECS a un grupo AS manualmente o mediante una acción de ajuste, pasa por los estados de grupo **Adding to AS group**, **Enabled** y **Removing from AS group**. A continuación, se elimina finalmente del grupo AS.

Si ha agregado un gancho de ciclo de vida al grupo AS, las estatuas de instancia cambian como se muestra en **Figura 4-2**. Cuando se produce un evento de escalamiento horizontal o de escalamiento horizontal en el grupo AS, las instancias requeridas se suspenden mediante el enlace de ciclo de vida y permanecen en el estado de espera hasta que finalice el período de tiempo de espera o se devuelva la llamada manualmente a las instancias. Puede realizar operaciones personalizadas en las instancias cuando estén en estado de espera. Por ejemplo, puede instalar o configurar software en una instancia antes de agregarla al grupo AS o descargar archivos de log de una instancia antes de quitarla.

Figura 4-2 Ciclo de vida de instancia



En las condiciones de activación 2, 4, 6 u 8, el sistema cambia automáticamente el estado de la instancia.

Tabla 4-2 Estado de instancia

Estado	Subestatos	Descripción del estado	Descripción del activador
Initial	No hay	La instancia no se ha agregado al grupo AS.	El estado de la instancia se cambia a <b>Adding to AS group</b> cuando se produce alguna de las siguientes situaciones: <ul style="list-style-type: none"> <li>● Aumenta manualmente el número esperado de instancias de un grupo AS.</li> <li>● El sistema agrega automáticamente instancias al grupo AS en una acción de ajuste.</li> <li>● Puede agregar instancias manualmente al grupo AS.</li> </ul>
Adding to AS group	Incorporación de una instancia.	Cuando se cumple la condición 1 de activación, AS agrega la instancia para ampliar la capacidad del grupo AS.	
Wait (Adding to AS group)	No hay	El gancho de ciclo de vida suspende la instancia que se está agregando al grupo AS y pone la instancia en estado de espera.	El estado de la instancia se cambia de <b>Wait (Adding to AS group)</b> a <b>Adding to AS group</b> cuando se realiza una de las siguientes operaciones: <ul style="list-style-type: none"> <li>● Se realiza la acción de devolución de llamada predeterminada.</li> <li>● Realice manualmente la acción de devolución de llamada.</li> </ul>

Estado	Subestatos	Descripción del estado	Descripción del activador
Adding to AS group	(Opcional) Asocia la instancia con un oyente de equilibrio de carga.	Cuando se cumple la condición de activación 3, el AS asocia la instancia con el oyente de equilibrado de carga.	
Enabled	No hay	La instancia se agrega al grupo AS y comienza a procesar el tráfico de servicio.	El estado de la instancia se cambia de <b>Enabled</b> a <b>Removing from AS group</b> cuando se produce cualquiera de las siguientes situaciones:
Removing from AS group	(Opcional) Disocia la instancia del oyente de equilibrio de carga.	Cuando se cumple la condición de activación 5, el grupo AS comienza a reducir recursos y disocia la instancia del oyente de equilibrado de carga.	<ul style="list-style-type: none"> <li>● Se reduce manualmente el número esperado de instancias de un grupo AS.</li> <li>● El sistema elimina automáticamente instancias en una acción de ajuste.</li> <li>● Una comprobación de estado muestra que la instancia no está en estado después de habilitarse y el sistema la elimina del grupo AS.</li> <li>● Eliminar manualmente una instancia de un grupo AS.</li> </ul>
Wait (Removing from AS group)	No hay	El gancho de ciclo de vida suspende la instancia que se está quitando del grupo AS y establece que la instancia esté en estado de espera.	El estado de la instancia se cambia de <b>Wait (Removing from AS group)</b> a <b>Removing from AS group</b> cuando se produce una de las siguientes situaciones:
Removing from AS group	Retire la instancia.	Cuando se cumple la condición de activación 7, el AS elimina la instancia del grupo AS.	<ul style="list-style-type: none"> <li>● Se realiza la acción de devolución de llamada predeterminada.</li> <li>● Realice manualmente la acción de devolución de llamada.</li> </ul>
Removed	No hay	El ciclo de vida de la instancia en el grupo AS finaliza.	No hay

Las instancias se agregan a un grupo AS manualmente o mediante una acción de ajuste. Luego, pasan por **Adding to AS group**, **Wait (Adding to AS group)**, **Adding to AS group**,



**Enabled, Removing from AS group, Wait (Removing from the AS group) y Removing from AS group** y finalmente se eliminan del grupo AS.

# 5 Restricciones

AS tiene las siguientes restricciones:

- Solo las aplicaciones que no tienen estado y se pueden escalar horizontalmente pueden ejecutarse en instancias de un grupo de AS.

## NOTA

- Un proceso o aplicación sin estado puede entenderse de forma aislada. No hay ningún conocimiento almacenado o referencia a transacciones pasadas. Cada transacción se realiza como si fuera desde cero por primera vez.  
Las instancias de ECS en las que se ejecutan aplicaciones sin estado no almacenan datos que deben persistir localmente.  
Piense en las transacciones apátridas como una máquina expendedora: una única solicitud y una respuesta.
- Las aplicaciones y procesos estatales, sin embargo, son aquellos a los que se puede volver una y otra vez. Se realizan en el contexto de transacciones anteriores y la transacción actual puede verse afectada por lo que ocurrió durante transacciones anteriores.  
Las instancias de ECS en las que se ejecutan aplicaciones con estado almacenan datos que deben persistir localmente.  
Las transacciones estatales se realizan repetidamente, como la banca en línea o el correo electrónico, que se realizan en el contexto de transacciones anteriores.
- AS puede liberar instancias de ECS en un grupo AS automáticamente, por lo que las instancias no se pueden usar para guardar información de estado de la aplicación (como estados de sesión) o datos relacionados (como datos de base de datos y registros). Si el estado de la aplicación o los datos relacionados deben guardarse, puede almacenar la información en servidores separados.
- AS no admite la expansión de capacidad ni la deducción de vCPUs y memoria de instancia.
- Los recursos de AS deben cumplir con los requisitos de cuotas enumerados en [Tabla 5-1](#).

**Tabla 5-1** Cuotas

Concepto	Descripción	Predeterminado
Grupo de AS	Número máximo de grupos AS por región por cuenta	10

<b>Concepto</b>	<b>Descripción</b>	<b>Predeterminado</b>
Configuración de AS	Número máximo de configuraciones AS por región por cuenta	100
Política de AS	Número máximo de políticas de AS por grupo de AS	10
instancia	Número máximo de instancias por grupo AS	300
política de escalado de ancho de banda	Número máximo de políticas de escalamiento de ancho de banda por región por cuenta	10

# 6 Región y AZ

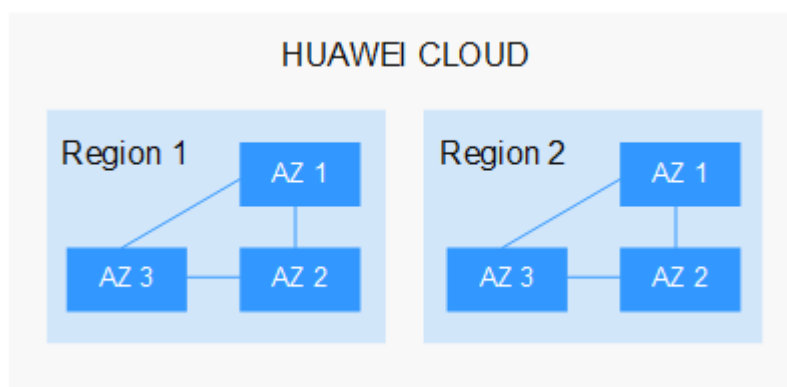
## Concepto

Una región y una zona de disponibilidad (AZ) identifican la ubicación de un centro de datos. Puede crear recursos en una región específica y AZ.

- Las regiones se dividen en función de la ubicación geográfica y la latencia de la red. Los servicios públicos, como Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Object Storage Service (OBS), Virtual Private Cloud (VPC), Elastic IP (EIP) y Image Management Service (IMS), se comparten dentro de la misma región. Las regiones se clasifican en regiones universales y regiones dedicadas. Una región universal proporciona servicios en la nube universales para los tenants estándares. Una región dedicada proporciona servicios específicos para tenants específicos.
- Una AZ contiene uno o más centros de datos físicos. Cada AZ cuenta con instalaciones independientes de electricidad, de refrigeración, de extinción de incendios y a prueba de humedad. Dentro de una AZ, los recursos de computación, red, almacenamiento y otros se dividen de forma lógica en múltiples clústeres. Las AZ dentro de una región están interconectadas usando fibras ópticas de alta velocidad, para soportar sistemas de alta disponibilidad entre las AZ.

**Figura 6-1** muestra la relación entre regiones y AZ.

**Figura 6-1** Las regiones y las AZ



Huawei Cloud ofrece servicios en muchas regiones de todo el mundo. Seleccione una región y AZ según los requisitos. Para obtener más información, consulte [Regiones globales de Huawei Cloud](#).

## Selección de una región

Al seleccionar una región, tenga en cuenta los siguientes factores:

- **Localización**  
Se recomienda seleccionar la región más cercana para una menor latencia de red y un acceso rápido. Las regiones dentro de China continental proporcionan la misma infraestructura, calidad de red BGP, así como operaciones de recursos y configuraciones. Por lo tanto, si sus usuarios objetivo están en China continental, no es necesario tener en cuenta las diferencias de latencia de la red al seleccionar una región.
  - Si sus usuarios objetivo se encuentran en Asia Pacífico (excepto China continental), seleccione la región **CN-Hong Kong, AP-Bangkok, or AP-Singapore**.
  - Si sus usuarios objetivo se encuentran en África, seleccione la región **AF-Johannesburg**.
  - Si sus usuarios objetivo están en América Latina, seleccione la región **LA-Santiago**.

### **NOTA**

La región **LA-Santiago** se encuentra en Chile.

- **Precio del recurso**  
Los precios de los recursos pueden variar en diferentes regiones. Para obtener más información, consulte [Detalles de precios del producto](#).

## Selección de una AZ

Al implementar recursos, tenga en cuenta los requisitos de las aplicaciones en cuanto a la recuperación ante desastres (DR) y la latencia de la red.

- Para una alta capacidad de DR, implemente recursos en diferentes AZ dentro de la misma región.
- Para una menor latencia de red, implemente recursos en la misma AZ.

## Regiones y endpoint

Antes de usar una API para llamar a recursos, especifique su región y endpoint.

# 7 Facturación

---

Puede usar AS de forma gratuita, pero las instancias de ECS creadas automáticamente en un grupo de AS se facturan de forma de pago por uso. Para obtener más información sobre los precios, consulte [Facturación de ECS](#). Los EIP utilizados por las instancias también se facturan. Para obtener detalles de precios, consulte [Facturación de EIP](#). Cuando el grupo AS escala, las instancias creadas automáticamente se eliminarán del grupo AS y se eliminarán. Después de la eliminación, estas instancias ya no se facturan. Las instancias agregadas manualmente se siguen facturando después de eliminarlas del grupo AS. Si no necesita estas instancias, anule su suscripción en la consola de ECS.

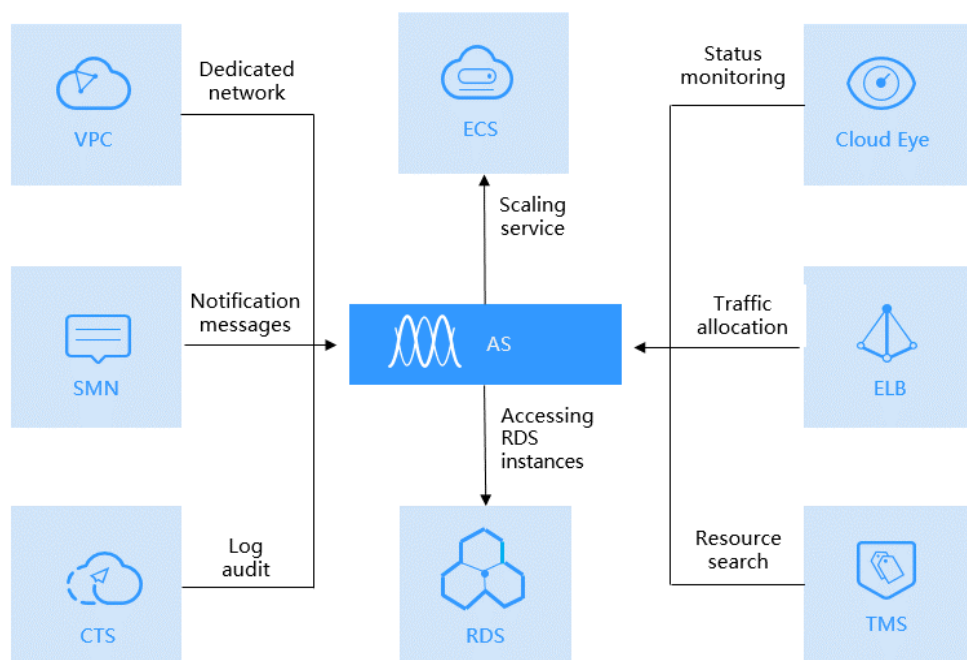
Por ejemplo, si se crean dos instancias cuando se escala un grupo AS, pero una hora más tarde, el grupo AS vuelve a escalar, las dos instancias se eliminan del grupo AS y se le facturará por esa hora de uso.

# 8 AS y otros servicios

AS puede trabajar con otros servicios en la nube para satisfacer sus requisitos para diferentes escenarios.

**Figura 8-1** muestra las relaciones entre AS y otros servicios.

**Figura 8-1** Relaciones entre AS y otros servicios



**Tabla 8-1** Servicios relacionados

Servicio	Descripción	Interacción	Referencia
Elastic Load Balance (ELB)	Después de configurar ELB, AS asocia automáticamente las instancias de ECS a un oyente del balanceador de carga al agregar ECS y las desvincula al eliminar las instancias.  Para que AS funcione con ELB, el grupo AS y el balanceador de carga deben estar en la misma VPC.	El AS distribuye el tráfico a todos los ECS de un grupo AS.	<a href="#">Adición de un balanceador de carga a un grupo de AS</a>
Cloud Eye	Si se configura una política activada por alarma, AS activa las acciones de ajuste cuando se cumple una condición de alarma especificada en Cloud Eye.	AS escala los recursos según el estado de la instancia de ECS monitoreado por Cloud Eye.	<a href="#">Métricas de AS</a>
ECS	Las instancias de ECS agregadas en una acción de ajuste se pueden gestionar y mantener en la consola de ECS.	AS ajusta automáticamente el número de instancias de ECS.	<a href="#">Dynamically Expanding Resources</a>
Virtual Private Cloud (VPC)	AS automatically adjusts the bandwidths of EIPs assigned in VPCs and also shared bandwidths.	AS automatically adjusts the bandwidth.	<a href="#">Creación de una política de ajuste de ancho de banda</a>



Servicio	Descripción	Interacción	Referencia
Simple Message Notification (SMN)	Si habilita el servicio SMN, el sistema le envía notificaciones sobre el estado de su grupo AS de manera oportuna.	Notificación con mensaje	<a href="#">Configuración de notificación para un grupo AS</a>
Cloud Trace Service (CTS)	Con CTS, puede grabar los registros de operaciones AS para ver, auditar y retroceder.	Auditoría de log	<a href="#">Grabación de operaciones de recursos AS</a>
Tag Management Service (TMS)	Si tiene varios recursos del mismo tipo, TMS le permite gestionar estos recursos más fácilmente.	Etiquetas	<a href="#">Adición de etiquetas a grupos e instancias de AS</a>
Relational Database Service (RDS)	Los requisitos previos para acceder directamente a una instancia de base de datos RDS desde una instancia escalada son los siguientes: <ul style="list-style-type: none"> <li>● La instancia escalada y la instancia de base de datos de RDS de destino deben estar en la misma VPC.</li> <li>● El grupo de seguridad debe permitir que la instancia escalada acceda a las instancias de base de datos de RDS.</li> </ul>	Las instancias escaladas pueden acceder a las instancias de base de datos de RDS.	<a href="#">Conexión a una instancia de base de datos RDS para MySQL a través de una red privada</a>

# 9 Gestión de permisos

---

Si necesita asignar diferentes permisos a los empleados de su empresa para acceder a sus recursos de AS, la Administración de Identidad y Acceso (IAM) es una buena opción para la gestión de permisos detallada. IAM proporciona autenticación de identidad, gestión de permisos y control de acceso, lo que le ayuda a proteger el acceso a sus recursos de Huawei Cloud.

Con IAM, puede crear usuarios de IAM y asignar permisos a los usuarios para controlar su acceso a recursos específicos. Por ejemplo, puede asignar permisos para permitir que algunos desarrolladores de software utilicen recursos de AS, pero no permitirles eliminar o realizar operaciones de alto riesgo en los recursos.

Si su cuenta de Huawei Cloud no necesita usuarios individuales de IAM para la gestión de permisos, omita esta sección.

IAM se puede utilizar de forma gratuita. Solo paga por los recursos de su cuenta. Para obtener más información acerca de IAM, consulte [Descripción general del servicio IAM](#).

## Permisos de AS

De forma predeterminada, los nuevos usuarios de IAM no tienen ningún permiso asignado. Debe agregarlos a uno o más grupos y adjuntar políticas o roles a estos grupos para que estos usuarios puedan heredar permisos de los grupos y realizar operaciones especificadas en los servicios en la nube.

Cuando concede permisos AS a un grupo de usuarios, establezca **Scope** en **Region-specific projects** y, a continuación, seleccione proyectos (por ejemplo, **ap-southeast-2** en la región **AP-Bangkok**) para que los permisos surtan efecto. Si selecciona **All projects**, los permisos surtirán efecto para el grupo de usuarios en todos los proyectos específicos de la región. Al acceder a AS, los usuarios deben cambiar a una región en la que se les haya autorizado a usar este servicio.

Puede conceder permisos a los usuarios mediante roles y políticas.

- Roles: Tipo de mecanismo de autorización de grano grueso que define permisos relacionados con las responsabilidades del usuario. Este mecanismo proporciona solo un número limitado de roles de nivel de servicio para la autorización. Al usar roles para conceder permisos, también debe asignar otros roles de los que dependen los permisos para que surtan efecto. Sin embargo, los roles no son una opción ideal para la autorización detallada y el control de acceso seguro.

- Políticas: Un tipo de mecanismo de autorización detallado que define los permisos necesarios para realizar operaciones en recursos de nube específicos bajo ciertas condiciones. Este mecanismo permite una autorización más flexible basada en políticas, cumpliendo los requisitos para un control de acceso seguro. Por ejemplo, puede conceder a los usuarios de AS solo los permisos para gestionar un determinado tipo de ECS. La mayoría de las políticas definen permisos basados en API. Para ver las acciones de API admitidas por AS, consulte [Políticas de permisos y acciones admitidas](#).

**Tabla 9-1** enumera todas las políticas de sistema admitidas por AS.

**Tabla 9-1** Permisos definidos por el sistema compatibles con AS

Nombre de la política	Descripción	Categoría	Dependencia
AutoScaling FullAccess	Todos los permisos de operación en todos los recursos de AS	Política definida por el sistema	No hay
AutoScaling ReadOnlyAccess	Permisos de sólo lectura en todos los recursos de AS	Política definida por el sistema	No hay
AutoScaling Administrator	Todos los permisos de operación en todos los recursos de AS	Rol del sistema	Los roles de <b>ELB Administrator</b> , <b>CES Administrator</b> , <b>Server Administrator</b> , y <b>Tenant Administrator</b> deben asignarse en el mismo proyecto.

**Tabla 9-2** enumera las operaciones comunes soportadas por cada política definida por el sistema de AS. Seleccione las políticas según sea necesario.

**Tabla 9-2** Operaciones comunes apoyadas por cada política definida por el sistema de AS

Operación	AutoScaling FullAccess	AutoScaling ReadOnlyAccess	AutoScaling Administrator
Creación de un grupo de AS	√	x	√
Modificación de un grupo de AS	√	x	√

Operación	AutoScaling FullAccess	AutoScaling ReadOnlyAccess	AutoScaling Administrator
Consulta de detalles sobre un grupo AS	√	√	√
Eliminación de un grupo de AS	√	x	√
Creación de una configuración de AS	√	x	√
Creación de una política de AS	√	x	√
Creación de una política de ajuste de ancho de banda	√	x	√

## Enlaces útiles

- [¿Qué es IAM?](#)
- [Creación de un usuario y concesión de permisos AS](#)
- [Políticas de permisos y acciones admitidas](#)

# 10 Conceptos básicos

---

## Grupo de AS

Un grupo AS consta de una colección de instancias de ECS que se aplican al mismo escenario. Es la base para habilitar o deshabilitar políticas de AS y realizar acciones de ajuste de capacidad.

## Configuración de AS

Una configuración de AS es una plantilla que especifica las especificaciones para las instancias de ECS que se agregarán a un grupo de AS. Las especificaciones incluyen el tipo ECS, vCPUs, memoria, imagen, modo de inicio de sesión y disco.

## Política de AS

Las políticas AS pueden desencadenar acciones de ajuste para ajustar el número de instancias en un grupo AS. Una política AS define la condición para activar una acción de ajuste y la operación que se realizará en una acción de ajuste. Cuando se cumple la condición de activación, el sistema activa automáticamente una acción de ajuste.

## Acción de ajuste de capacidad

Una acción de ajuste agrega instancias a o elimina instancias de un grupo AS. Garantiza que el número esperado de instancias se estén ejecutando en el grupo AS mediante la adición o eliminación de instancias cuando se cumple la condición de activación, lo que mejora la estabilidad del sistema.

## Período de recuperación

Para evitar que una política basada en alarmas se active repetidamente por el mismo evento, configure un período de enfriamiento. Un período de tiempo de reutilización especifica durante cuánto tiempo se rechazará una acción de escalada activada por alarma después de que se haya completado una acción de escalada anterior. Este período de tiempo de reutilización no se aplica a las acciones de escala programadas o periódicas.

Por ejemplo, si estableces el período de tiempo de reutilización en 300 segundos (5 minutos), y hay una acción de ajuste programada para 10:32, pero una acción de ajuste anterior se completó a las 10:30, Las acciones de ajuste activadas por alarma se denegarán durante el período de enfriamiento de 10:30 a 10:35, pero la acción de ajuste programada se activará a

las 10:32. Si la acción de ajuste programada termina a las 10:36, un nuevo período de tiempo de reutilización comienza a las 10:36 y termina a las 10:41.

## **Ajuste de ancho de banda**

AS ajusta automáticamente el ancho de banda según las políticas de ajuste que haya configurado. AS solo puede ajustar los anchos de banda de los EIP y compartir anchos de banda que se facturan sobre una base de pago por uso.

# 11 Historial de revisiones

---

Lanzado en	Novedades
2021-10-30	Modificado el siguiente contenido: Agregada la sección "Gestión de permisos".
2020-10-19	Modificado el siguiente contenido: Agregada la sección "Métodos de acceso".
2019-09-30	Esta edición es el segundo lanzamiento oficial.
2018-11-19	Esta edición es el primer lanzamiento oficial.