

Auto Scaling

Início rápido

Edição 01
Data 2018-11-19



Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2023. Todos os direitos reservados.

Nenhuma parte deste documento pode ser reproduzida ou transmitida em qualquer forma ou por qualquer meio sem consentimento prévio por escrito da Huawei Cloud Computing Technologies Co., Ltd.

Marcas registadas e permissões



HUAWEI e outras marcas registadas da Huawei são marcas registadas da Huawei Technologies Co., Ltd.

Todas as outras marcas registadas e os nomes registados mencionados neste documento são propriedade dos seus respectivos detentores.

Aviso

Os produtos, os serviços e as funcionalidades adquiridos são estipulados pelo contrato estabelecido entre a Huawei Cloud e o cliente. Os produtos, os serviços e as funcionalidades descritos neste documento, no todo ou em parte, podem não estar dentro do âmbito de aquisição ou do âmbito de uso. Salvo especificação em contrário no contrato, todas as declarações, informações e recomendações neste documento são fornecidas "TAL COMO ESTÃO" sem garantias ou representações de qualquer tipo, sejam expressas ou implícitas.

As informações contidas neste documento estão sujeitas a alterações sem aviso prévio. Foram feitos todos os esforços na preparação deste documento para assegurar a exatidão do conteúdo, mas todas as declarações, informações e recomendações contidas neste documento não constituem uma garantia de qualquer tipo, expressa ou implícita.

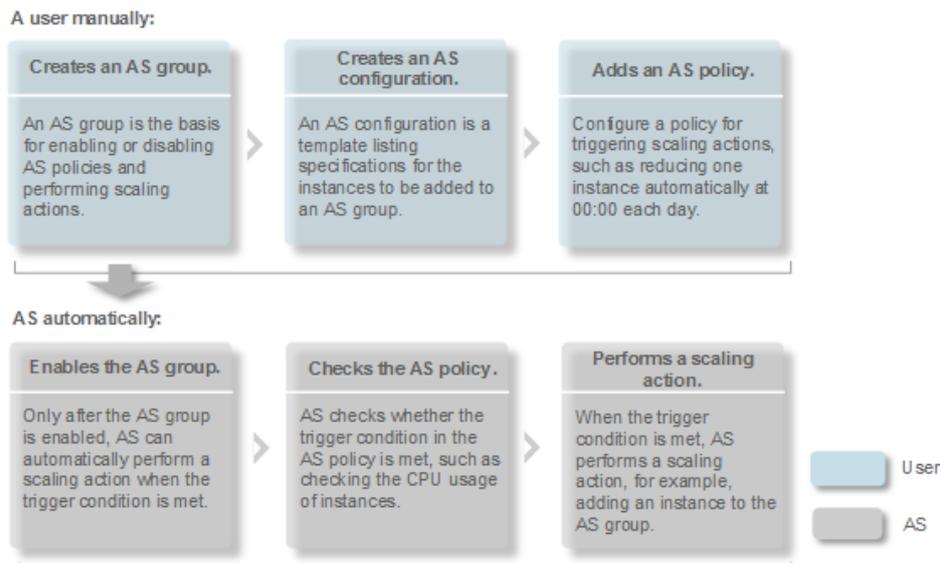
Índice

| | |
|---|----------|
| 1 Processo baseado em assistente de uso de AS..... | 1 |
| 2 Criação rápida de um grupo de AS..... | 2 |

1 Processo baseado em assistente de uso de AS

Figura 1-1 ilustra o processo baseado em assistente de usar AS.

Figura 1-1 Processo baseado em assistente de usar AS



2 Criação rápida de um grupo de AS

Se você estiver usando AS pela primeira vez, seguir o processo baseado em assistente é uma maneira fácil de criar um grupo de AS, uma configuração de AS e uma política de AS.

Pré-requisitos

- Você criou as VPCs, sub-redes, grupos de segurança e balanceadores de carga necessários.
- Você obteve os pares de chaves para fazer logon nas instâncias adicionadas por uma ação de dimensionamento se a autenticação de chave for usada.

Procedimento

1. Faça logon no console de gerenciamento.
2. Em **Compute**, clique em **Auto Scaling**. No painel de navegação à esquerda, escolha **Instance Scaling**.
3. Clique em **Create AS Group**.
4. Defina informações básicas sobre o grupo de AS, como **Name**, **Max. Instances**, **Min. Instances** e **Expected Instances**. [Tabela 2-1](#) lista os parâmetros.

Tabela 2-1 Parâmetros do grupo de AS

| Parâmetro | Descrição | Exemplo de valor |
|-----------|---|------------------|
| Region | Uma região é onde o grupo de AS é implementado. Recursos em diferentes regiões não podem se comunicar uns com os outros através de redes internas. Para menor latência de rede e acesso mais rápido aos seus recursos, selecione a região mais próxima dos seus usuários de destino. | N/A |

| Parâmetro | Descrição | Exemplo de valor |
|-------------------------|---|------------------|
| AZ | <p>Uma AZ é um local físico onde os recursos usam fontes de alimentação e redes independentes. As AZs são fisicamente isoladas, mas interconectadas por meio de uma rede interna.</p> <ul style="list-style-type: none"> ● Se você precisar de alta disponibilidade, compre servidores em AZs diferentes. ● Se você precisar de baixa latência de rede, compre servidores na mesma AZ. | N/A |
| Multi-AZ Scaling Policy | <p>Este parâmetro pode ser definido como Load-balanced ou Sequenced.</p> <ul style="list-style-type: none"> ● Load-balanced: ao expandir um grupo de AS, o sistema preferencialmente distribui as instâncias de ECS uniformemente entre as AZs usadas pelo grupo de AS. Se falhar na AZ de destino, selecionará automaticamente outra AZ com base na política sequenciada. ● Sequenced: ao expandir ECSs em um grupo de AS, o sistema seleciona a AZ de destino com base na ordem em que as AZs são selecionadas. <p>NOTA Esse parâmetro precisa ser configurado quando duas ou mais AZs forem selecionadas.</p> | Load-balanced |
| Name | <p>Especifica o nome do grupo de AS a ser criado. O nome pode conter até 64 caracteres. Pode incluir letras, dígitos, sublinhados (_) e hifens (-).</p> | N/A |
| Max. Instances | Especifica o número máximo de instâncias do ECS em um grupo de AS. | 1 |
| Expected Instances | <p>Especifica o número esperado de instâncias do ECS em um grupo de AS.</p> <p>Depois que um grupo de AS é criado, você pode alterar esse valor, o que acionará uma ação de dimensionamento.</p> | 0 |
| Min. Instances | Especifica o número mínimo de instâncias do ECS em um grupo de AS. | 0 |
| VPC | <p>Fornecer uma rede para suas instâncias do ECS. Todas as instâncias do ECS no grupo do AS são implementadas nesta VPC.</p> | N/A |

| Parâmetro | Descrição | Exemplo de valor |
|----------------|---|------------------|
| Subnet | <p>Você pode selecionar até cinco sub-redes. O grupo de AS vincula automaticamente todas as NICs aos ECSs criados. A primeira sub-rede é usada pela NIC primária de uma instância do ECS por padrão e outras sub-redes são usadas pelas NICs de extensão da instância.</p> | N/A |
| Load Balancing | <p>Este parâmetro é opcional. Um balanceador de carga distribui automaticamente o tráfego em todas as instâncias de um grupo de AS para equilibrar sua carga de serviço. Ele melhora a tolerância a falhas de suas aplicações e expande os recursos de serviços de aplicações.</p> <p>NOTA</p> <ul style="list-style-type: none"> ● Até seis balanceadores de carga podem ser adicionados a um grupo de AS. ● Depois que vários balanceadores de carga são adicionados a um grupo de AS, vários serviços podem ser ouvidos simultaneamente, melhorando assim a escalabilidade. Se ELB health check for selecionada para Health Check Method, quando qualquer um dos ouvintes detectar que uma instância não íntegra, o AS substituirá a instância defeituosa por uma funcional. <p>Se você selecionar Elastic load balancer, configure os seguintes parâmetros:</p> <ul style="list-style-type: none"> ● Load Balancer ● Backend ECS Group ● Backend Port: especifica a porta na qual um ECS de back-end escuta o tráfego. ● Weight: determina a parte das solicitações que um ECS de back-end processa em comparação com outros ECSs de back-end adicionados ao mesmo ouvinte. <p>Para obter mais informações sobre balanceamento de carga, consulte <i>Guia de usuário do Elastic Load Balance</i>.</p> | N/A |

| Parâmetro | Descrição | Exemplo de valor |
|-------------------------|--|------------------|
| Instance Removal Policy | <p>Especifica a prioridade para remover instâncias de um grupo de AS. Se as condições especificadas forem atendidas, as ações de dimensionamento serão acionadas para remover instâncias. O AS suporta as seguintes políticas de remoção de instâncias:</p> <ul style="list-style-type: none"> ● Oldest instance created from oldest AS configuration: a instância mais antiga criada a partir da configuração mais antiga é removida primeiro do grupo de AS. ● Newest instance created from oldest AS configuration: a instância mais recente criada a partir da configuração mais antiga é removida primeiro do grupo de AS. ● Oldest instance: a instância mais antiga é removida primeiro do grupo de AS. ● Newest instance: a instância mais recente é removida primeiro do grupo de AS. <p>NOTA</p> <ul style="list-style-type: none"> ● A remoção de instâncias garantirá preferencialmente que as instâncias restantes tenham balanceamento de carga nas AZs. ● As instâncias do ECS adicionadas manualmente são as últimas a serem removidas. Se o AS remover uma instância adicionada manualmente, ele removerá apenas a instância do grupo de AS. Ele não exclui a instância. Se várias instâncias adicionadas manualmente precisarem ser removidas, o AS preferencialmente removerá primeiro a instância adicionada mais cedo. | N/A |
| EIP | <p>Se EIP tiver sido selecionado em uma configuração de AS para um grupo de AS, um EIP será automaticamente vinculado à instância do ECS adicionada ao grupo de AS. Se você selecionar Release, o EIP vinculado a uma instância será liberado quando a instância for removida do grupo de AS. Caso contrário, o sistema desvincula o EIP da instância, mas não o libera quando a instância é removida do grupo de AS.</p> | N/A |

| Parâmetro | Descrição | Exemplo de valor |
|-----------------------|--|------------------|
| Health Check Method | <p>Quando uma verificação de integridade detecta uma instância de ECS não íntegra, o AS substitui a instância defeituosa por uma nova. Existem dois métodos de verificação de integridade:</p> <ul style="list-style-type: none"> ● ECS health check: verifica o status de integridade da instância do ECS. Se uma instância for interrompida ou excluída, ela é considerada não íntegra. Este método é selecionado por padrão. Usando esse método, o grupo de AS avalia periodicamente o status de execução de cada instância com base nos resultados da verificação de integridade. Se os resultados da verificação de integridade mostrarem que uma instância não é íntegra, o AS removerá a instância do grupo de AS. ● ELB health check: determina o status de execução da instância do ECS usando um ouvinte de balanceamento de carga. Esse método de verificação de integridade só estará disponível se o grupo de AS usar um ouvinte de balanceamento de carga. Se um ouvinte de balanceamento de carga detectar que uma instância do ECS não está íntegra, o AS removerá a instância do grupo de AS. | N/A |
| Health Check Interval | Especifica o período de tempo entre as verificações de integridade. Você pode definir um intervalo de verificação de integridade, como 10 segundos, 1 minuto, 5 minutos, 15 minutos, 1 hora ou 3 horas, com base nos requisitos do site. | 5 minutes |
| Enterprise Project | <p>Especifica o projeto empresarial ao qual o grupo de AS pertence. Se um projeto empresarial é configurado para um grupo de AS, os ECSS criados neste grupo de AS também pertencem a este projeto empresarial. Se você não especificar um projeto empresarial, o projeto empresarial default será usado.</p> <p>NOTA</p> <ul style="list-style-type: none"> ● Valor default indica o projeto empresarial padrão. Os recursos que não estão alocados a nenhum projeto empresarial na sua conta são exibidos no projeto empresarial padrão. ● O projeto empresarial é uma versão atualizada do IAM. Ele aloca e gerencia recursos de diferentes projetos. | N/A |
| Advanced Settings | <p>Configure notificações.</p> <p>Você pode selecionar Do not configure ou Configure now.</p> | N/A |

| Parâmetro | Descrição | Exemplo de valor |
|--------------|--|------------------|
| Notification | <p>Os resultados das ações de dimensionamento são enviados a você com base nas funções fornecidas pelo serviço Simple Message Notification (SMN).</p> <ul style="list-style-type: none"> ● Notification Conditions: quando pelo menos uma das condições a seguir for atendida, o SMN enviará uma notificação para você: <ul style="list-style-type: none"> - Criação de instância foi bem-sucedida - Remoção da instância foi bem-sucedida - Erros ocorrem em um grupo de AS - Falha na criação da instância - Falha na remoção da instância ● Send Notification To: selecione um tópico existente. Para obter detalhes sobre como criar um tópico, consulte <i>Guia de usuário do Simple Message Notification</i>. | N/A |
| Tag | <p>Se você tiver muitos recursos do mesmo tipo, poderá usar tags para gerenciar seus recursos. Você pode identificar recursos especificados rapidamente usando as tags alocadas a eles.</p> <p>Cada tag contém uma chave e um valor. Você pode especificar a chave e o valor para cada tag.</p> <ul style="list-style-type: none"> ● Chave <ul style="list-style-type: none"> - A chave deve ser especificada. - A chave deve ser exclusiva para o grupo de AS. - A chave pode incluir até 36 caracteres. Ela não pode conter caracteres ASCII não imprimíveis (0 - 31) ou os seguintes caracteres: =* <> \, / ● Valor <ul style="list-style-type: none"> - O valor é opcional. - Uma chave pode ter apenas um valor. - O valor pode incluir até 43 caracteres. Ela não pode conter caracteres ASCII não imprimíveis (0 - 31) ou os seguintes caracteres: =* <> \, / | N/A |

5. Clique em **Next**.
6. Na página exibida, você pode usar uma configuração do AS existente ou criar uma configuração do AS.

7. Clique em **Next**.
8. (Opcional) Adicione uma política do AS a um grupo do AS.
Na página exibida, clique em **Add AS Policy**.
Configure os parâmetros necessários, como **Policy Type**, **Scaling Action** e **Cooldown Period**.

 **NOTA**

- Se uma ação de dimensionamento for acionada por uma política de AS, o período de resfriamento será o que estiver configurado para essa política de AS.
 - Se uma ação de dimensionamento for acionada alterando manualmente o número esperado de instâncias ou por outras ações, o período de resfriamento será o que estiver configurado para o grupo de AS. O período de resfriamento padrão é de 300 segundos.
9. Clique em **Create Now**.
 10. Verifique o grupo de AS, a configuração de AS e as informações de política de AS.
Clique em **Submit**.
 11. Confirme o resultado da criação e volte para a página **AS Groups** conforme solicitado.
Depois que o grupo de AS é criado, seu status muda para **Enabled**.