

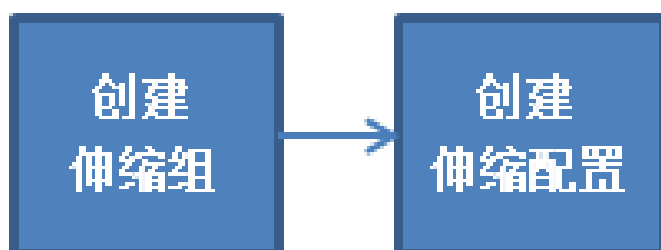
弹性伸缩

快速入门

快速入门

两步创建伸缩方案

创建完整的弹性伸缩方案主要包含以下 2 步：



登录弹性伸缩控制台，单击 **整体方案创建管理**。



创建伸缩组

伸缩组是具有相同应用场景的 ECS 实例的集合。您可以定义伸缩组内 ECS 实例数的最大值、最小值及其相关联的负载均衡实例和 RDS 实例。

- 填写页面列出的各选项。
- 弹性伸缩将根据您设置的 **伸缩最小实例数**，在伸缩组内创建相同个数的实例。

选择 **负载均衡** 实例时，指定的负载均衡实例所有的监听端口必须开启健康检查。

1. 伸缩组基本信息 伸缩组是具有相同应用场景的ECS实例的集合。伸缩组定义了组内ECS实例数的最大值、最小值及其相关联的SLB实例和RDS实例等属性。

*地域：北京
更改地域将清空选择的负载均衡、数据库实例和实例配置。

*伸缩组名称：ESS示例
名称为2-40个字符，以大小写字母、数字或中文开头，可包含“.”、“_”或“-”

*伸缩最大实例数（台）：2
最大值100

*伸缩最小实例数（台）：2
请填写正整数

*默认冷却时间（秒）：300
最小为0，必须为整数

释放策略：执行释放 最早伸缩配置对应的实例 在结果中再筛选 最早创建的实例 执行释放

负载均衡：148a0b6a978-cn-beijing-btc-a01 SLB选定以后将不支持修改

数据库（个）：1 选择数据库 数据库选定以后将不支持修改

取消 下一步

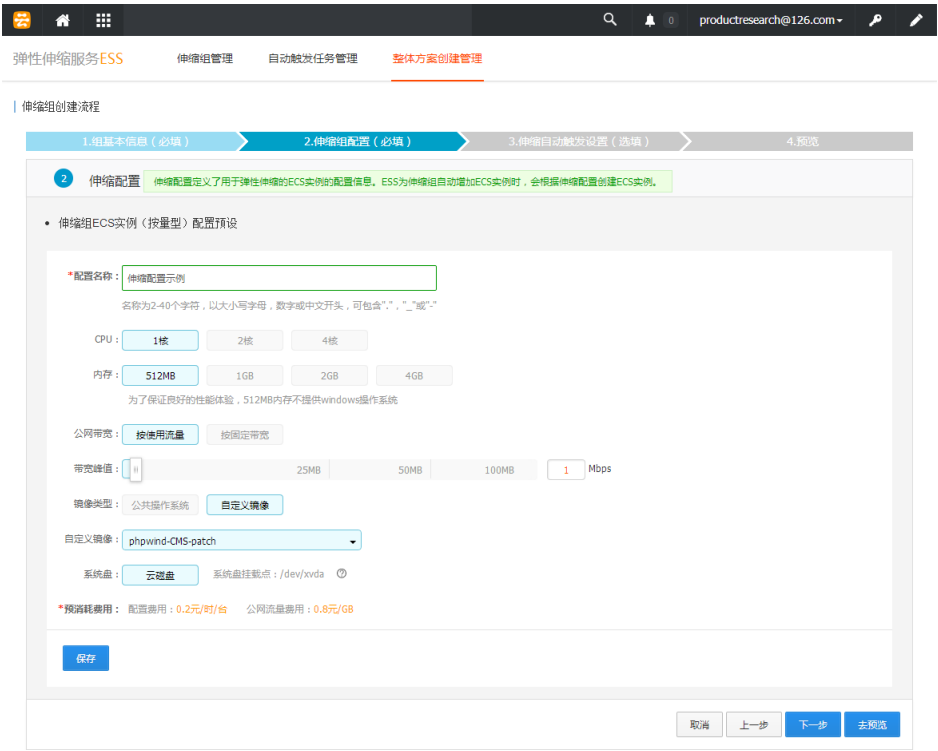
选择完成后，单击 **下一步**。

创建伸缩配置

根据伸缩配置中您设置的实例具体规格，弹性伸缩自动创建 ECS 实例并添加进伸缩组。

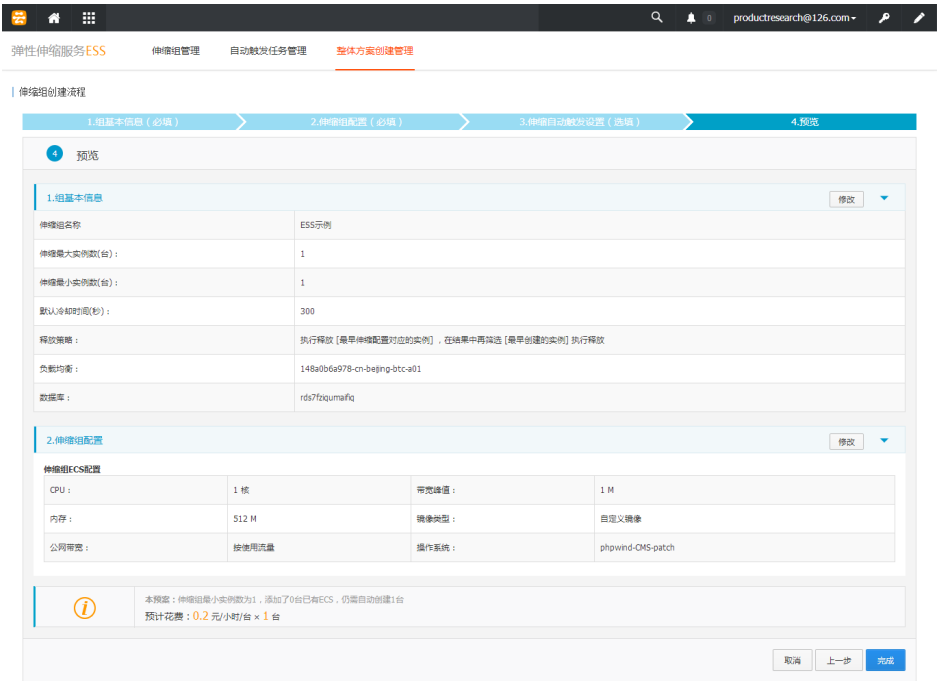
- 为页面列出的各选项做出选择。
- 在选择**镜像类型**时，如果您需要实现自动启动 Web 服务器、自动下载代码和脚本等功能，请选择**自定义镜像**。

单击 **保存** > **去预览**。



预览及完成方案创建

预览页面展示前两步设置的伸缩组、伸缩配置信息，以及该方案在创建时预计花费的费用。费用仅包含 ECS 实例的费用，弹性伸缩服务免费。



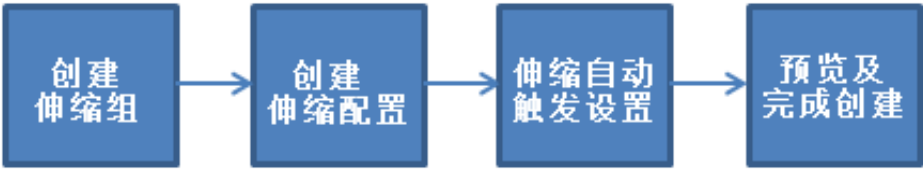
单击 完成，展示整个方案的创建结果。

说明：弹性伸缩会自动创建伸缩配置中您设置的最小实例数量的ECS 实例，并自动将实例加入指定的负载均衡实例，将实例的 IP 加入指定的 RDS 访问白名单当中。



四步创建完整伸缩方案

创建包含自动触发设置的整体伸缩方案，包含以下 4 步。



登录弹性伸缩控制台，单击 导航栏中的**整体方案创建管理**。

1. 创建伸缩组

参见两步创建伸缩方案。



2. 创建伸缩配置

参见两步创建伸缩方案。

点击保存之后，点击下一步，进入伸缩自动触发设置（选填）页面。



3. 伸缩自动触发设置

伸缩自动触发有 2 种模式：

- 定时触发自动伸缩任务。
- 报警触发自动伸缩任务。

定时触发自动伸缩任务

弹性伸缩根据您设置的执行时间和重复周期，自动触发伸缩任务。如果未设置重复周期，则仅触发一次伸缩任务。



报警触发自动伸缩任务

基于云监控性能指标（如 CPU、内存利用率），自动触发伸缩任务。此功能需要在 ECS 的镜像里安装最新版本的云监控 Agent。



4. 预览及完成创建

预览页面将展示前几步中您的设置，以及该方案的预计费用。您只需要支付 ECS 实例产生的费用，弹性伸缩服务免费。



单击 **完成**，将展示整个方案的创建结果。

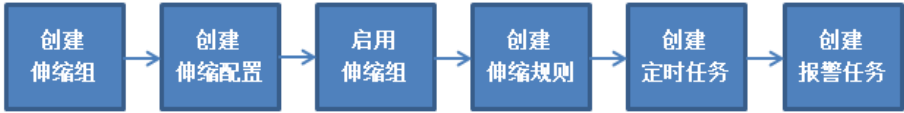


API快速入门

流程介绍

本文介绍如何通过 Open API 创建和配置包含定时、动态、自定义、固定模式的整体伸缩方案。

创建完整的弹性伸缩方案，您需要通过以下图中所示步骤来完成（其中前三步为简易的伸缩方案）：



创建伸缩组（CreateScalingGroup），配置伸缩资源的最小值（MinSize）、最大值（MaxSize）及需要关联的负载均衡实例和 RDS 实例。

创建伸缩配置（CreateScalingConfiguration），指定需要弹性伸缩的 ECS 实例的相关属性，如 ImageID、InstanceType 等。

以第二步创建的伸缩配置启用伸缩组（EnableScalingGroup）。

创建伸缩规则（CreateScalingRule），如 **加 N 台 ECS 实例** 的伸缩规则。

创建伸缩任务（CreateScheduledTask），如创建 12：00 触发第四步伸缩规则的定时任务。

创建报警任务（云监控 API PutAlarmRule），如创建 CPU 大于等于 80% 则增加一台 ECS 实例的报警任务。

创建伸缩组

创建伸缩组（CreateScalingGroup），配置伸缩资源的最小值（MinSize）、最大值（MaxSize）及需要关联的负载均衡实例和 RDS 实例。

请求示例

```
http://ess.aliyuncs.com/?Action=CreateScalingGroup
&RegionId=cn-qingdao
&MaxSize=20
&MinSize=2
&LoadBalancerId=147b46d767c-cn-qingdao-cm5-a01
&DBInstanceId.1=rdszzzyunybaeu
&DBInstanceId.2=rdsia3u3yia3u3y
&<公共请求参数>
```

返回示例

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

创建伸缩配置

创建伸缩配置（CreateScalingConfiguration），指定需要弹性伸缩的 ECS 实例的相关属性，如ImageID、InstanceType 等。请求时，需要指定第一步返回的 ScalingGroupId。

请求示例

```
http://ess.aliyuncs.com/?Action=CreateScalingConfiguration
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbcrBT1
&SecurityGroupId=sg-280ih3w4b
&ImageId=centos6u5_64_20G_aliaegis_20140703.vhd
&InstanceType=ecs.t1.xsmall
&<公共请求参数>
```

返回示例

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

启用伸缩组

以第二步创建的伸缩配置启用伸缩组（EnableScalingGroup），还可以在此步加入一台既有的 ECS 实例。

请求示例

```
http://ess.aliyuncs.com/?Action=EnableScalingGroup
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbcrBT1
&ActiveScalingConfigurationId=eOs27Kb0oXvQcUYjEGelJqUy
&InstanceId.1=i-283vvytn
&<公共请求参数>
```

返回示例

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

创建伸缩规则

本文是API请求CreateScalingRule（创建伸缩规则）的示例。在请求时，需要指定ScalingGroupId。

请求示例

```
http://ess.aliyuncs.com/?Action=CreateScalingRule  
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbcrBT1  
&AdjustmentType=QuantityChangeInCapacity  
&AdjustmentValue=1  
&<公共请求参数>
```

返回示例

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

创建定时任务

本文是API请求CreateScheduledTask（创建伸缩任务）的示例。在请求时，需要指定 ScalingRuleAri。

请求示例

```
http://ess.aliyuncs.com/?Action=CreateScheduledTask  
&RegionId=cn-qingdao
```

```
&LaunchTime=2014-08-17T12:00Z
&RecurrenceType=Daily
&RecurrenceValue=1
&RecurrenceEndTime=2014-09-17T16:55Z
&ScheduledAction=ari:acs:ess:cn-qingdao:1344371:scalingrule/eMKWG8SRNb9dBLAjweNI1Ik
&<公共请求参数>
```

返回示例

```
<CreateScheduledTaskResponse>
<ScheduledTaskId>edRtShc57WGXdt8TIPbrjsnV</ScheduledTaskId>
<RequestId>0F02D931-2B12-44D7-A0E9-39925C13D15E</RequestId>
</CreateScheduledTaskResponse>
```

产品限制

弹性伸缩的 ECS 实例中部署的应用需要是无状态、可横向扩展的。

由于弹性伸缩会自动释放 ECS 实例，所以用于弹性伸缩的 ECS 实例不可以保存应用的状态信息（如 session）和相关数据（如数据库、日志等）。如果应用中需要保存状态信息，可以考虑把状态信息保存到独立的状态服务器、数据库（如 RDS）及集中日志存储（如 SLS）。

目前，弹性伸缩不会将自动扩展出来的实例自动添加到 开放缓存（OCS）访问白名单中,需要您自行添加。

弹性伸缩目前不支持纵向扩展，即弹性伸缩暂时无法自动升降 ECS 的 CPU、内存和带宽。