
Amazon EC2 Auto Scaling

用户指南



Amazon EC2 Auto Scaling: 用户指南

Copyright © 2019 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

什么是 Amazon EC2 Auto Scaling ?	1
Auto Scaling 组件	1
入门	1
访问 Amazon EC2 Auto Scaling	2
Amazon EC2 Auto Scaling 定价	2
PCI DSS 合规性	2
相关服务	2
Auto Scaling 的优势	3
示例：覆盖可变需求	3
示例：Web 应用程序架构	4
示例：在可用区之间分配实例	5
Auto Scaling 生命周期	6
扩展	7
已投入使用的实例	7
缩小	7
附加实例	8
分离实例	8
生命周期挂钩	8
进入和退出备用状态	8
Amazon EC2 Auto Scaling 限制	8
设置	10
注册 AWS	10
准备使用 Amazon EC2	10
入门	11
步骤 1：创建启动模板	11
步骤 2：创建 Auto Scaling 组	13
步骤 3：验证 Auto Scaling 组	14
(可选) 终止您的 Auto Scaling 组中的实例	15
步骤 4：(可选) 删除扩展基础设施	15
教程：设置具有扩展和负载均衡功能的应用程序	17
先决条件	17
配置扩展和负载均衡功能 (控制台)	17
创建或选择启动配置	17
创建或选择启动模板	18
创建 Auto Scaling 组	19
(可选) 验证您的负载均衡器是否已附加到您的 Auto Scaling 组	20
配置扩展和负载均衡功能 (AWS CLI)	20
创建启动配置	20
创建启动模板	21
创建包含负载均衡器的 Auto Scaling 组	21
清除您的 AWS 资源	21
启动模板	22
为 Auto Scaling 组创建启动模板	22
将启动配置复制到启动模板	26
使用启动模板替换启动配置	27
启动配置	28
创建启动配置	28
使用 EC2 实例创建启动配置	29
使用 EC2 实例创建启动配置	30
从实例创建启动配置以及覆盖块储存设备 (AWS CLI)	31
创建启动配置和覆盖实例类型 (AWS CLI)	32
更改启动配置	33
在 VPC 中启动 Auto Scaling 实例	33
默认 VPC	34

VPC 中的 IP 寻址	34
实例部署租期	34
将 EC2-Classic 实例链接到 VPC	35
Auto Scaling 组	37
使用多种实例类型和购买选项	37
分配策略	38
控制按需实例的比例	39
Spot 实例的最佳实践	40
先决条件	40
使用多个购买选项创建 Auto Scaling 组	40
使用启动模板创建组	44
使用启动配置创建组	45
使用 EC2 实例创建组	46
使用 EC2 实例创建 Auto Scaling 组 (控制台)	47
使用 EC2 实例创建 Auto Scaling 组 (AWS CLI)	47
使用启动向导创建组	48
标记 Auto Scaling 组和实例	49
标签限制	49
标记生命周期	50
对 Auto Scaling 组添加或修改标签	50
删除标签	52
将负载均衡器与 Auto Scaling 组结合使用	53
Elastic Load Balancing 类型	53
附加负载均衡器	54
添加 ELB 运行状况检查	56
添加可用区	57
在 Auto Scaling 组中启动 Spot 实例	59
合并 Auto Scaling 组	59
合并区域 (AWS CLI)	60
删除 Auto Scaling 基础设施	61
删除 Auto Scaling 组	61
(可选) 删除启动配置	61
(可选) 删除启动模板	62
(可选) 删除负载均衡器	62
(可选) 删除 CloudWatch 警报	63
扩展组	64
扩展选项	64
保持 Auto Scaling 组的大小	65
Amazon EC2 Auto Scaling 的手动扩展	65
更改您的 Auto Scaling 组的大小 (控制台)	65
更改您的 Auto Scaling 组的大小 (AWS CLI)	66
将 EC2 实例附加到 Auto Scaling 组	67
从 Auto Scaling 组中分离 EC2 实例	71
计划的扩展	74
注意事项	74
创建和管理计划操作 (控制台)	74
创建和管理计划操作 (AWS CLI)	75
动态扩展	76
扩展策略类型	76
多个扩展策略	77
目标跟踪扩展策略	77
简单扩展和步进扩展策略	81
向现有的 Auto Scaling 组添加扩展策略	88
基于 Amazon SQS 进行扩展	89
删除扩展策略	92
扩展冷却时间	93
示例：冷却时间	93

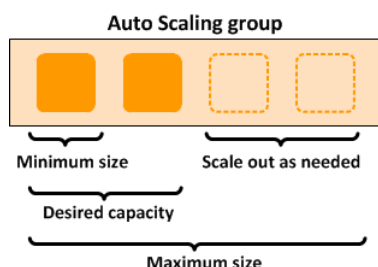
默认冷却时间	94
特定于扩展的冷却时间	94
冷却时间和多个实例	95
冷却时间和生命周期挂钩	95
Auto Scaling 实例终止	95
默认终止策略	95
自定义终止策略	97
实例保护	98
生命周期挂钩	100
生命周期挂钩的工作方式	100
使用生命周期挂钩时的注意事项	101
准备通知	102
添加生命周期挂钩	105
完成生命周期挂钩	105
测试通知	106
临时删除实例	106
备用状态的工作方式	107
处于备用状态的实例的运行状况	107
使用临时删除实例 (控制台)	107
临时删除实例 (AWS CLI)	108
暂停扩展	110
扩展流程	110
选择暂停	111
暂停和恢复扩展进程 (控制台)	113
暂停和恢复扩展 (AWS CLI)	113
监控您的 Auto Scaling 实例和组	115
运行状况检查	115
实例运行状况	115
确定实例运行状况	116
运行状况检查宽限期	116
替换运行状况不佳的实例	116
自定义运行状况检查	117
Amazon CloudWatch 指标	117
Auto Scaling 组指标	118
Auto Scaling 组指标的维度	118
启用 Auto Scaling 组指标	118
为 Auto Scaling 实例配置监控	119
查看 CloudWatch 指标	120
创建 Amazon CloudWatch 警报	121
Amazon CloudWatch Events	122
Auto Scaling 事件	122
创建 Lambda 函数	126
将事件路由到 Lambda 函数	126
Amazon SNS 通知	127
SNS 通知	128
配置 Amazon SNS	128
配置 Auto Scaling 组以发送通知	129
测试通知配置	130
验证是否收到扩展事件的通知	130
删除通知配置	131
AWS CloudTrail 日志记录	132
CloudTrail 中的 Amazon EC2 Auto Scaling 信息	132
了解 Amazon EC2 Auto Scaling 日志文件条目	132
控制对您的资源的访问权	134
在策略中指定操作	134
其他 IAM 权限	134
指定资源	135

在策略中指定条件	136
预定义的 AWS 托管策略	137
客户托管策略示例	137
示例：限制能够传递哪些服务相关角色（使用 PassRole）	137
示例：需要启动模板	138
示例：创建并管理启动配置	139
示例：创建并管理 Auto Scaling 组和扩展策略	140
示例：使用标签控制访问	141
示例：更改 Auto Scaling 组的容量	142
服务相关角色	143
AWSServiceRoleForAutoScaling 授予的权限	143
创建服务相关角色（自动）	144
创建服务相关角色（手动）	144
编辑服务相关角色	145
删除服务相关角色	145
Amazon EC2 Auto Scaling 服务相关角色支持的区域	145
与加密卷结合使用时必需的 CMK 密钥策略	145
示例：允许访问 CMK 的 CMK 密钥策略部分	145
示例：允许跨账户访问 CMK 的 CMK 密钥策略部分	146
在 Amazon EC2 实例上运行的应用程序的 IAM 角色	147
先决条件	148
创建启动配置	148
创建启动模板	148
故障排除	149
检索错误消息	149
实例启动失败	151
安全组 <该安全组的名称> 不存在。启动 EC2 实例失败。	151
密钥对 <与您的 EC2 实例相关联的密钥对> 不存在。启动 EC2 实例失败。	151
当前不支持请求的配置。	152
AutoScalingGroup <Auto Scaling 组名称> 未找到。	152
不再支持请求的可用区。请重新尝试请求...	152
请求的实例类型（<实例类型>）在请求的可用区（<实例可用区>）中不受支持...	152
您没有订阅此项服务。请参阅 https://aws.amazon.com/ 。	152
设备名称上载无效。启动 EC2 实例失败。	152
用于参数 virtualName 的值（<与实例存储设备相关联的名称>）无效...	153
实例存储 AMI 不支持 EBS 块存储设备映射。	153
置放群组可能无法与类型为“m1.large”的实例一起使用。启动 EC2 实例失败。	153
Client.InternalError：客户端在启动时出错。	153
AMI 问题	154
AMI ID <您的 AMI 的 ID> 不存在。启动 EC2 实例失败。	154
AMI <AMI ID> 正在等待，无法运行。启动 EC2 实例失败。	155
用于参数 virtualName 的值（<ami ID>）无效。	155
请求的实例类型的架构（i386）与用于 ami-6622f00f（x86_64）的清单中的架构不匹配。启动 EC2 实例失败。	155
负载均衡器问题	155
找不到负载均衡器 <您的启动环境>。验证负载均衡器配置失败。	156
名为 <负载均衡器名称> 的活动负载均衡器不存在。更新负载均衡器配置失败。	156
EC2 实例 <实例 ID> 不在 VPC 内。更新负载均衡器配置失败。	156
EC2 实例 <实例 ID> 在 VPC 中。更新负载均衡器配置失败。	156
请求中包含的安全令牌无效。验证负载均衡器配置失败。	156
容量限制	156
您请求的可用区（<请求的可用区>）中当前没有足够的 <实例类型> 容量...	157
已运行 <实例数量> 个实例。启动 EC2 实例失败。	157
资源	158
文档历史记录	159

什么是 Amazon EC2 Auto Scaling ?

Amazon EC2 Auto Scaling 助您确保拥有适量的 Amazon EC2 实例，用于处理您的应用程序负载。您可创建 EC2 实例的集合，称为 Auto Scaling 组。您可以指定每个 Auto Scaling 组中最少的实例数量，Amazon EC2 Auto Scaling 会确保您的组中的实例永远不会低于这个数量。您可以指定每个 Auto Scaling 组中最大的实例数量，Amazon EC2 Auto Scaling 会确保您的组中的实例永远不会高于这个数量。如果您在创建组的时候或在创建组之后的任何时候指定了所需容量，Amazon EC2 Auto Scaling 会确保您的组一直具有此数量的实例。如果您指定了扩展策略，则 Amazon EC2 Auto Scaling 可以在您的应用程序的需求增加或降低时启动或终止实例。




例如，以下 Auto Scaling 组的最小大小为 1 个实例，所需的容量为 2 个实例，最大大小为 4 个实例。您制定的扩展策略是按照您指定的条件，在最大最小实例数范围内调整实例的数量。



有关 Amazon EC2 Auto Scaling 优点的更多信息，请参阅 [Auto Scaling 的优势 \(p. 3\)](#)。

Auto Scaling 组件

下表描述 Amazon EC2 Auto Scaling 的关键组件。

	<p>组</p> <p>您的 EC2 实例整理到组 中，从而当作一个逻辑单位进行扩展和管理。当您创建一个组时，您可以指定其中 EC2 实例的最小数量、最大数量以及所需数量。有关更多信息，请参阅 Auto Scaling 组 (p. 37)。</p>
	<p>配置模板</p> <p>您的组将启动模板 或启动配置 作为其 EC2 实例的配置模板。您可以为实例指定一些信息，例如，AMI ID、实例类型、密钥对、安全组 and 块储存设备映射。有关更多信息，请参阅 启动模板 (p. 22) 和 启动配置 (p. 28)。</p>
	<p>扩展选项</p> <p>Amazon EC2 Auto Scaling 提供了多种扩展 Auto Scaling 组的方式。例如，您可以将组配置为在发生指定条件时（动态扩展）或根据时间表进行扩展。有关更多信息，请参阅扩展选项 (p. 64)。</p>

入门

如果您刚刚接触 Amazon EC2 Auto Scaling，建议您在开始工作之前查看 [Auto Scaling 生命周期 \(p. 6\)](#)。

若要开始使用，请学完[Amazon EC2 Auto Scaling 入门 \(p. 11\)](#)教程，以创建一个 Auto Scaling 组并了解它在该组中的实例终止时如何进行响应。如果您已有 EC2 实例正在运行，则可用一个现有 EC2 实例创建 Auto Scaling 组，并可随时从该组中删除实例。

访问 Amazon EC2 Auto Scaling

如果您已注册 AWS 账户，可以通过如下方式访问 Amazon EC2 Auto Scaling 控制台：登录 AWS 管理控制台 并从控制台主页选择 EC2，然后从导航窗格中选择 Auto Scaling Groups (Auto Scaling 组)。

您也可以使用 [Amazon EC2 Auto Scaling API](#) 访问 Amazon EC2 Auto Scaling。Amazon EC2 Auto Scaling 提供了查询 API。这些请求属于 HTTP 或 HTTPS 请求，需要使用 HTTP 动词 GET 或 POST 以及一个名为 Action 的查询参数。有关 Amazon EC2 Auto Scaling 的 API 操作的更多信息，请参阅 Amazon EC2 Auto Scaling API 参考 中的 [操作](#)。

如果您倾向于使用特定语言的 API 而非通过 HTTP 或 HTTPS 提交请求来构建应用程序，AWS 为软件开发人员提供了库文件、示例代码、教程和其他资源。这些库文件提供可自动执行任务的基本功能，例如以加密方式对请求签名、重试请求和处理错误响应，因此您可以更轻松地上手。有关更多信息，请参阅 [AWS SDKs and Tools](#)。

如果倾向于使用命令行界面，您可使用以下选项：

AWS 命令行接口 (CLI)

提供大量 AWS 产品的相关命令，同时被 Windows、macOS 和 Linux 支持。要了解其用法，请参阅 [AWS Command Line Interface 用户指南](#)。有关更多信息，请参阅 AWS CLI Command Reference 中的 [自动扩展](#)。

适用于 Windows PowerShell 的 AWS 工具

为在 PowerShell 环境中编写脚本的用户提供大量 AWS 产品的相关命令。要开始使用，请参阅 [适用于 Windows PowerShell 的 AWS 工具 用户指南](#)。有关更多信息，请参阅 [适用于 PowerShell 的 AWS 工具 Cmdlet Reference](#)。

有关用于访问 AWS 的凭证的信息，请参阅 Amazon Web Services 一般参考 中的 [AWS 安全凭证](#)。有关用于调用 Amazon EC2 Auto Scaling 的区域和终端节点的信息，请访问 AWS General Reference 中的 [AWS 区域和终端节点](#)。

Amazon EC2 Auto Scaling 定价

Amazon EC2 Auto Scaling 不产生额外费用，因此可方便地试用它并了解它如何使您的 AWS 架构获益。

PCI DSS 合规性

Auto Scaling 支持由商家或服务提供商处理、存储和传输信用卡数据，而且已经验证符合支付卡行业 (PCI) 数据安全标准 (DSS)。有关 PCI DSS 的更多信息，包括如何请求 AWS PCI Compliance Package 的副本，请参阅 [PCI DSS 第 1 级](#)。

相关服务

要为您的应用程序的所有可扩展 AWS 资源配置自动扩展，请使用 AWS Auto Scaling。借助 AWS Auto Scaling，您还可以简化为 Auto Scaling 组定义动态扩展策略的流程，并使用预测扩展功能在提前预测到流量变化时扩展您的 Amazon EC2 容量。有关更多信息，请参阅 [AWS Auto Scaling 用户指南](#)。

使用 Elastic Load Balancing 在您 Auto Scaling 组的多个实例之间自动分配应用程序的传入流量。有关更多信息，请参阅 [Elastic Load Balancing 用户指南](#)。

要监控您的实例和 Amazon EBS 卷的基本统计数据，可使用 Amazon CloudWatch。有关更多信息，请参阅 [Amazon CloudWatch 用户指南](#)。

要监控您的账户对 Amazon EC2 Auto Scaling API 进行的调用，请使用 AWS CloudTrail。记录的数据包括 AWS 管理控制台、命令行工具和其他服务进行的调用。有关更多信息，请参阅 [AWS CloudTrail User Guide](#)。

Auto Scaling 的优势

将 Amazon EC2 Auto Scaling 添加到应用程序架构是一种最大限度利用 AWS 云的方法。当您使用 Amazon EC2 Auto Scaling 时，您的应用程序将获得以下优势：

- 提高容错能力。Amazon EC2 Auto Scaling 可以检测到实例何时运行状况不佳并终止实例，然后启动新实例以替换它。您还可以配置 Amazon EC2 Auto Scaling 以使用多个可用区。如果一个可用区变得不可用，则 Amazon EC2 Auto Scaling 可以在另一个可用区中启动实例以进行弥补。
- 提高了可用性。Amazon EC2 Auto Scaling 有助于确保应用程序始终具有合适的容量以满足当前的流量需求。
- 加强成本管理。Amazon EC2 Auto Scaling 可以根据需要动态地增加或降低容量。由于您仅为使用的 EC2 实例付费，您可以在需要的时候启动实例，并在不需要的时候终止实例以节约成本。

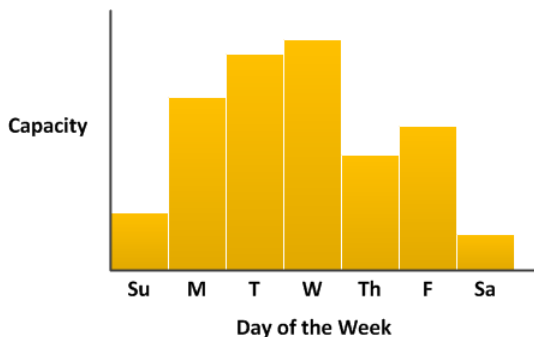
目录

- [示例：覆盖可变需求 \(p. 3\)](#)
- [示例：Web 应用程序架构 \(p. 4\)](#)
- [示例：在可用区之间分配实例 \(p. 5\)](#)

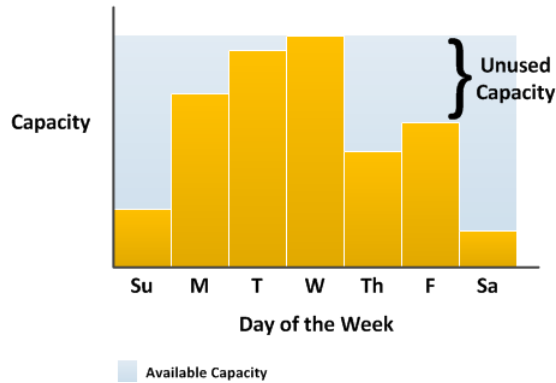
示例：覆盖可变需求

为了说明 Amazon EC2 Auto Scaling 的一些优势，请设想一个在 AWS 上运行的基本 Web 应用程序。此应用程序允许员工搜索可用于开会的会议室。每周开始和结束时段，此应用程序的使用率最低。每周中期，有更多的员工安排会议，因此对此应用程序的需求会显著提高。

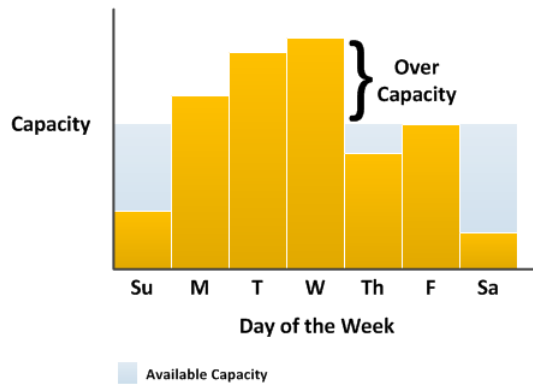
下图显示此应用程序的容量在一周中的使用情况。



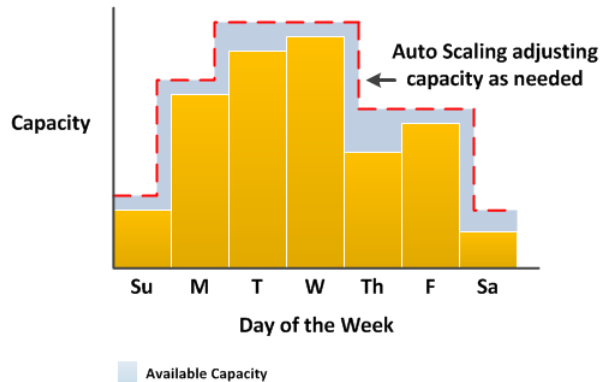
按照传统做法，可通过两种方式对这些容量变化做好规划。第一种选择是添加足够多的服务器，以便应用程序始终具有足够的容量来满足需求。但是，这种做法的缺点是应用程序在某些天并不需要这么多容量。额外容量闲置不用，并且实际上提高了使应用程序保持运行的成本。



第二种选择是采用处理应用程序平均需求所需的容量。这种做法成本更低，因为不用购买仅仅偶尔使用的设备。然而，这样做的风险是：当对应用程序的需求超过其容量时，可能造成糟糕的客户体验。

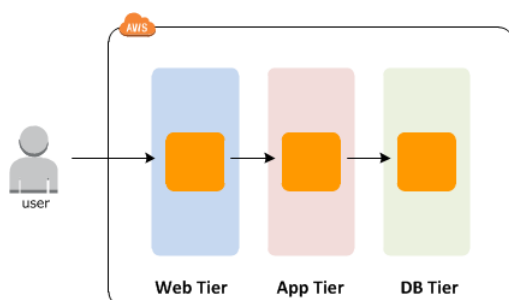


通过向此应用程序添加 Amazon EC2 Auto Scaling，您就有了第三种选择。您可以仅在需要时才向应用程序添加新实例，并在不再需要这些实例时终止它们。由于 Amazon EC2 Auto Scaling 使用 EC2 实例，因此您只需在使用时为使用的实例付费。您现在有了一个具有成本效益的架构，可在尽量减少支出的同时提供最佳客户体验。

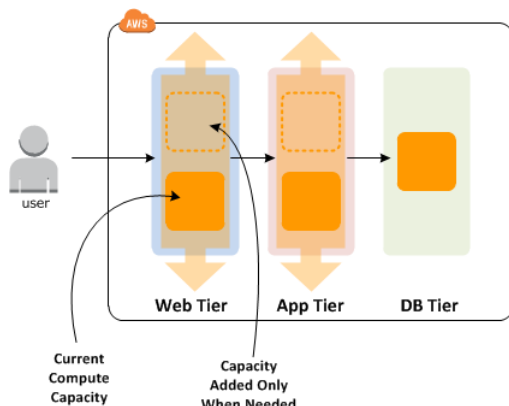


示例：Web 应用程序架构

在常见的 Web 应用程序场景中，您同时运行应用程序的多个副本来满足客户流量。将在相同的 EC2 实例（云服务器）上托管应用程序的多个副本，每个实例都可以处理客户请求。



Amazon EC2 Auto Scaling 代表您管理 EC2 实例的启动和终止。您定义一组条件（如 Amazon CloudWatch 警报），用于确定 Auto Scaling 组何时启动或终止 EC2 实例。将 Auto Scaling 组添加到网络架构有助于提高应用程序的可用性和容错能力。



您可以根据需要创建任意数量的 Auto Scaling 组。例如，您可以为每个层创建一个 Auto Scaling 组。

要在您的 Auto Scaling 组的各实例之间分配流量，可在您的架构中引入一个负载均衡器。有关更多信息，请参阅 [将负载均衡器与 Auto Scaling 组结合使用](#) (p. 53)。

示例：在可用区之间分配实例

AWS 资源（如 EC2 实例）放置在高度可用的数据中心。为了提供额外的扩展性和可靠性，这些数据中心位于不同的物理位置。地区是指大型、分布范围广泛的地理位置。每个区域包含多个不同的位置（称为可用区），它们旨在与其他可用区中的故障隔离。相同地区中的不同可用区之间通过廉价、低延迟的网络相连。有关更多信息，请参阅 Amazon Web Services 一般参考 中的 [区域和终端节点 Amazon EC2 Auto Scaling](#)。

在使用 Amazon EC2 Auto Scaling 时，您可以使 Auto Scaling 组跨区域中的多个可用区以利用地理冗余的安全性和可靠性。当一个可用区运行状况不佳或无法使用时，Auto Scaling 将在不受影响的可用区中启动新实例。当运行状况不佳的可用区恢复运行状况时，Auto Scaling 会自动在所有指定的可用区中重新均匀分配应用程序实例。

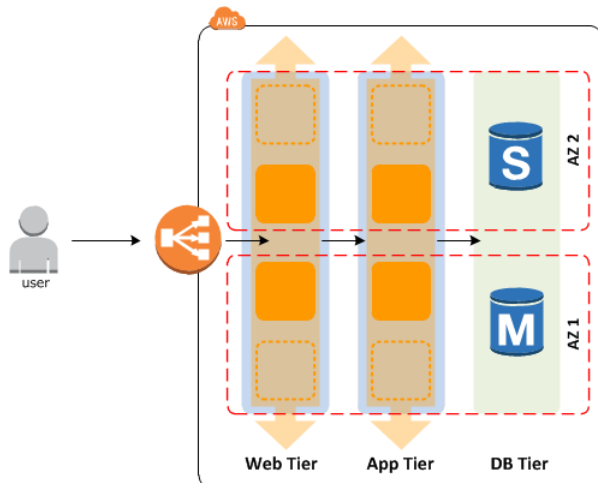
Auto Scaling 组可以包含来自同一区域中的一个或多个可用区的 EC2 实例。但是，Auto Scaling 组不能跨多个区域。

对于 VPC 中的 Auto Scaling 组，在子网中启动 EC2 实例。在您创建或更新 Auto Scaling 组时，要为您的 EC2 实例选择子网。您可为每个可用区选择一个或多个子网。有关更多信息，请参阅 Amazon VPC 用户指南中的 [VPC 和子网](#)。

实例分配

Amazon EC2 Auto Scaling 尝试在为您的 Auto Scaling 组启用的可用区之间均匀分配实例。Amazon EC2 Auto Scaling 通过尝试在实例最少的可用区中启动新实例来执行此操作。但是，如果尝试失败，Amazon

EC2 Auto Scaling 将尝试在另一可用区中启动实例，直到成功。对于 VPC 中的 Auto Scaling 组，如果可用区内有多个子网，Amazon EC2 Auto Scaling 将从可用区中随机选择一个子网。



再平衡活动

在某些操作发生后，您的 Auto Scaling 组可能会在不同可用区之间变得不平衡。Amazon EC2 Auto Scaling 将通过重新平衡可用区来进行补偿。以下操作可能导致重新平衡活动：

- 更改组的可用区。
- 您明确地终止或断开实例，组变得不平衡。
- 之前没有足够容量的可用区已恢复，并提供了额外容量。

再平衡时，Amazon EC2 Auto Scaling 在终止旧实例之前启动新实例，所以再平衡不会损害应用程序的性能或可用性。

因为 Amazon EC2 Auto Scaling 尝试在终止旧实例之前启动新实例，所以达到或接近指定的最大容量可能会阻碍或完全停止再平衡活动。为避免此问题，在再平衡活动期间，系统可以暂时超出某组的指定最大容量的 10%（或超出 1 个实例，以较大者为准）。仅当该组达到或接近最大容量，并需要重新平衡时，才可超出容量限制；此类情况的原因是用户请求重新分区，或者是为了弥补区域可用性问题的。该超出状态仅持续重新平衡该组所需的时间（通常为几分钟）。

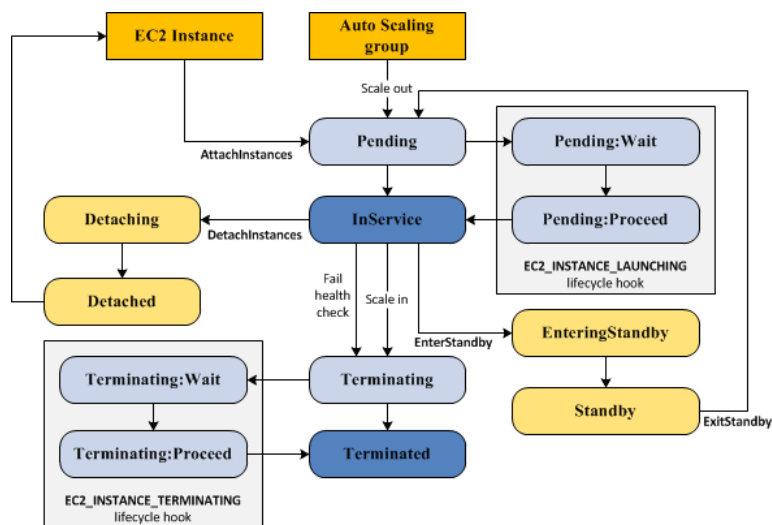
Auto Scaling 生命周期

Auto Scaling 组中的 EC2 实例具有的路径或生命周期不同于其他 EC2 实例中的路径或生命周期。生命周期从 Auto Scaling 组启动实例并将其投入使用时开始。生命周期在您终止实例或 Auto Scaling 组禁用实例并将其终止时结束。

Note

一旦启动实例，您就需要为实例付费，包括尚未将实例投入使用的時間。

下图阐释了 Amazon EC2 Auto Scaling 生命周期内的实例状态之间的过渡。



扩展

以下扩展事件指示 Auto Scaling 组启动 EC2 实例并将其附加到组：

- 手动增大组的大小。有关更多信息，请参阅 [Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)。
- 您创建一个扩展策略来自动根据指定的所需增量来增大组的大小。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的动态扩展 \(p. 76\)](#)。
- 您可以通过安排在某个特定时间增大组的大小来设置扩展。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的计划扩展 \(p. 74\)](#)。

在发生扩展事件时，Auto Scaling 组将使用分配的启动配置来启动所需数目的 EC2 实例。这些实例最初处于 Pending 状态。如果您向 Auto Scaling 组添加生命周期挂钩，则可在此处执行自定义操作。有关更多信息，请参阅 [生命周期挂钩 \(p. 8\)](#)。

在每个实例完全配置并通过 Amazon EC2 运行状况检查后，该实例将附加到 Auto Scaling 组并进入 InService 状态。针对 Auto Scaling 组的所需容量对实例进行计数。

已投入使用的实例

实例将保持 InService 状态，直至出现下列情况之一：

- 发生缩减事件，并且 Amazon EC2 Auto Scaling 选择终止此实例来减小 Auto Scaling 组的大小。有关更多信息，请参阅 [控制在缩小过程中终止哪些 Auto Scaling 实例 \(p. 95\)](#)。
- 将实例置于 Standby 状态。有关更多信息，请参阅 [进入和退出备用状态 \(p. 8\)](#)。
- 您从 Auto Scaling 组分离实例。有关更多信息，请参阅 [分离实例 \(p. 8\)](#)。
- 实例未通过所需数目的运行状况检查，因此将从 Auto Scaling 组中删除实例、终止实例和替换实例。有关更多信息，请参阅 [Auto Scaling 实例的运行状况检查 \(p. 115\)](#)。

缩小

以下缩小事件指示 Auto Scaling 组从组中分离 EC2 实例并将其终止：

- 手动减小组的大小。有关更多信息，请参阅 [Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)。
- 您创建一个扩展策略，自动根据指定的所需减少量来减小组的大小。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的动态扩展 \(p. 76\)](#)。

- 您可以通过安排在某个特定时间减小组的大小来设置扩展。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的计划扩展 \(p. 74\)](#)。

您必须为您创建的每个扩展事件创建一个相应的缩减事件。这有助于确保分配给您的应用程序的资源与对这些资源的需求尽可能相符。

发生缩小事件时，Auto Scaling 组分离一个或多个实例。Auto Scaling 组使用其终止策略来确定要终止的实例。正在从 Auto Scaling 组中分离和关闭的实例将进入 `Terminating` 状态，且无法重新将其投入使用。如果您向 Auto Scaling 组添加生命周期挂钩，则可在此处执行自定义操作。最后，实例将完全终止并进入 `Terminated` 状态。

附加实例

您可以将符合特定条件的正在运行的 EC2 实例附加到您的 Auto Scaling 组。在附加实例后，将该实例作为 Auto Scaling 组的一部分进行管理。

有关更多信息，请参阅 [将 EC2 实例附加到 Auto Scaling 组 \(p. 67\)](#)。

分离实例

您可以从 Auto Scaling 组中分离实例。分离实例后，您可以独立于 Auto Scaling 组管理实例或者将实例附加到其他 Auto Scaling 组。

有关更多信息，请参阅 [从 Auto Scaling 组中分离 EC2 实例 \(p. 71\)](#)。

生命周期挂钩

您可以将生命周期挂钩添加到 Auto Scaling 组，以便在实例启动或终止时执行自定义操作。

当 Amazon EC2 Auto Scaling 响应扩展事件时，它将启动一个或多个实例。这些实例最初处于 `Pending` 状态。如果您已将一个 `autoscaling:EC2_INSTANCE_LAUNCHING` 生命周期挂钩添加到您的 Auto Scaling 组，则实例将从 `Pending` 状态转换为 `Pending:Wait` 状态。完成生命周期操作后，实例将进入 `Pending:Proceed` 状态。在完全配置实例后，实例将附加到 Auto Scaling 组并进入 `InService` 状态。

当 Amazon EC2 Auto Scaling 响应缩减事件时，它将终止一个或多个实例。这些实例将从 Auto Scaling 组中分离并进入 `Terminating` 状态。如果您已将一个 `autoscaling:EC2_INSTANCE_TERMINATING` 生命周期挂钩添加到您的 Auto Scaling 组，则实例将从 `Terminating` 状态转换为 `Terminating:Wait` 状态。完成生命周期操作后，实例将进入 `Terminating:Proceed` 状态。在完全终止实例后，实例将进入 `Terminated` 状态。

有关更多信息，请参阅 [Amazon EC2 Auto Scaling 生命周期挂钩 \(p. 100\)](#)。

进入和退出备用状态

可以将任何处于 `InService` 状态的实例置于 `Standby` 状态。这使您能够终止对实例的使用，排查实例的问题或对实例进行更改，然后重新将实例投入使用。

处于 `Standby` 状态的实例继续由 Auto Scaling 组管理。但是，在将这些实例重新投入使用前，它们不是您的应用程序的有效部分。

有关更多信息，请参阅 [临时从 Auto Scaling 组中删除实例 \(p. 106\)](#)。

Amazon EC2 Auto Scaling 限制

您的 AWS 账户存在以下与 Amazon EC2 Auto Scaling 相关的限制。有关其他服务的限制的信息，请参阅 Amazon Web Services 一般参考 中的 [AWS 服务限制](#)。

默认限制

- 每个区域的启动配置数：200
- 每个区域的 Auto Scaling 组数：200

要查看您的账户的当前限制，请使用 Amazon EC2 控制台的限制页面或 [describe-account-limits](#) (AWS CLI) 命令。要请求增大限制，请使用 [Auto Scaling 限制表](#)。

Auto Scaling 组限制

- 每个 Auto Scaling 组的扩展策略数：50
- 每个 Auto Scaling 组的计划操作数：125
- 每个 Auto Scaling 组的生命周期挂钩数：50
- 每个 Auto Scaling 组的 SNS 主题数：10
- 每个 Auto Scaling 组的 Classic 负载均衡器数：50
- 每个 Auto Scaling 组的目标组数：50

扩展策略限制

- 每个扩展策略的分步调整数：20

Auto Scaling API 限制

- 您每次可以将 [AttachInstances](#)、[DetachInstances](#)、[EnterStandby](#) 和 [ExitStandby](#) 与最多 20 个实例 ID 一起使用。
- 您每次可以将 [AttachLoadBalancers](#) 和 [DetachLoadBalancers](#) 与最多 10 个负载均衡器一起使用。
- 您每次可以将 [AttachLoadBalancerTargetGroups](#) 和 [DetachLoadBalancerTargetGroups](#) 与最多 10 个目标组一起使用。

设置 Amazon EC2 Auto Scaling

开始使用 Amazon EC2 Auto Scaling 前，请完成以下任务。

任务

- [注册 AWS \(p. 10\)](#)
- [准备使用 Amazon EC2 \(p. 10\)](#)

注册 AWS

创建 AWS 账户时，我们会自动为所有 AWS 服务注册您的账户。您只需为使用的服务付费。除了要为 EC2 实例支付费用，您使用 Amazon EC2 Auto Scaling 无需任何额外费用。

如果您没有 AWS 账户，请按照以下方式注册 AWS。

如需注册 AWS 账户

1. 打开 <https://portal.aws.amazon.com/billing/signup>。
2. 按照屏幕上的说明进行操作。

在注册时，您将接到一通电话，要求您使用电话键盘输入一个验证码。

注册过程完成后，AWS 会向您发送一封确认电子邮件。

准备使用 Amazon EC2

如果您以前未使用过 Amazon EC2，请完成 Amazon EC2 文档中所述的任务。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[使用 Amazon EC2 进行设置](#)或 Amazon EC2 用户指南（适用于 Windows 实例）中的[使用 Amazon EC2 进行设置](#)。

Amazon EC2 Auto Scaling 入门

在使用 Amazon EC2 Auto Scaling 时，您必须从特定的构建块入手。本教程将分步为您介绍为 Amazon EC2 Auto Scaling 设置基本基础设施的过程。

在创建用于应用程序的 Auto Scaling 组之前，请全面检查应用程序在 AWS 云中运行时的情况。记录以下内容：

- 启动和配置服务器需要多长时间。
- 哪些指标与应用程序的性能关系最密切。
- Auto Scaling 组应跨多少个可用区。
- 可以使用哪些现有的资源，如安全组或亚马逊系统映像 (AMI)。
- 您是希望进行扩展以增加或减少容量，还是只希望确保始终运行特定数量的服务器？请记住，Amazon EC2 Auto Scaling 可以同时实现这两个目标。

您越了解您的应用程序，Auto Scaling 架构的效率就越高。

以下说明用于一个配置模板，该模板定义 EC2 实例，创建 Auto Scaling 组以保持固定实例数（即使实例已不能正常运行）以及选择删除该基本基础设施。

本教程假定您熟悉 EC2 实例的启动并已创建密钥对和安全组。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[设置 Amazon EC2](#)。

任务

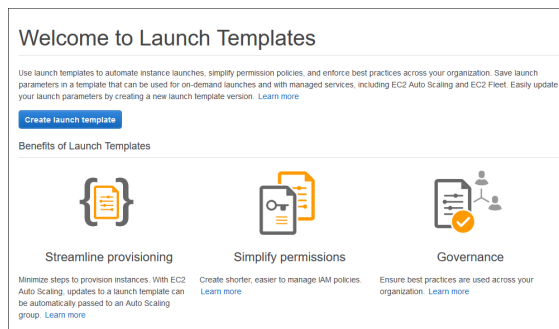
- [步骤 1：创建启动模板](#) (p. 11)
- [步骤 2：创建 Auto Scaling 组](#) (p. 13)
- [步骤 3：验证 Auto Scaling 组](#) (p. 14)
- [步骤 4：\(可选\) 删除扩展基础设施](#) (p. 15)

步骤 1：创建启动模板

在此步骤中，您将创建一个启动模板，它指定 Amazon EC2 Auto Scaling 为您创建的 EC2 实例的类型。包含一些信息，例如将使用的 Amazon 系统映像 (AMI) 的 ID、实例类型、密钥对、安全组和块储存设备映射。

为 Auto Scaling 组创建启动模板

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏中，选择 AWS 区域。您创建的 Amazon EC2 Auto Scaling 资源与您所指定的区域相关联。
3. 在导航窗格中，选择启动模板。
4. 选择创建启动模板。



5. 选择创建新模板。对于启动模板名称，输入一个名称（例如，`my_template`）。
6. 对于模板版本描述，输入启动模板初始版本的描述，以帮助您记住此模板的用途（例如，`test launch template for an Auto Scaling group`）。
7. 对于 AMI ID，请从 Quick Start (快速启动) 列表表中选择一个 Amazon Linux 2 (HVM) 版本。Amazon 系统映像 (AMI) 用作实例的基本配置模板。
8. 对于实例类型，选择与您指定的 AMI 兼容的硬件配置。请注意，免费套餐 Linux 服务器是一个 `t2.micro` 实例。

Note

如果您的账户不到 12 个月，您可以在特定使用限制下免费使用 `t2.micro` 实例。有关更多信息，请参阅 [AWS 免费套餐](#)。

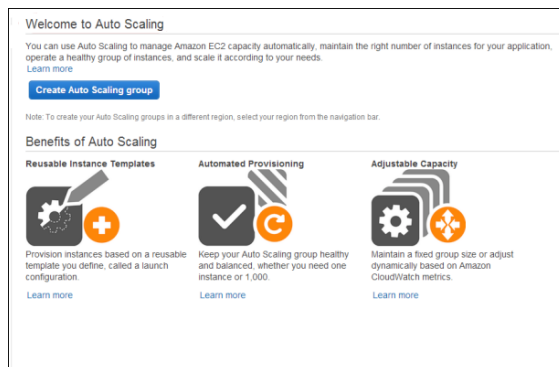
9. (可选) 对于密钥对名称，键入在连接到您的实例时使用的密钥对的名称。
10. (可选) 对于 Network type (网络类型)，选择 VPC。
11. 跳过安全组以在下一步中配置安全组。指定了网络接口时，安全组必须属于该接口。
12. 对于网络接口，执行以下操作以指定主网络接口：
 - a. 选择添加网络接口。
 - b. (可选) 要将公有 IP 地址分配给非默认 VPC 中的实例，对于自动分配公有 IP，请选择 启用。这将允许您的实例与 Internet 或 AWS 中的其他服务通信。
 - c. 对于安全组 ID，指定网络接口的安全组。
 - d. 对于终止时删除，请选择在 Auto Scaling 组缩减并终止网络接口附加到的实例时是否删除网络接口。
13. (可选) 对于存储 (卷)，除了您指定的 AMI 所指定的卷以外，还可以指定要附加到实例的卷。
14. (可选) 对于实例标签，指定一个或多个要与实例和卷关联的标签。
15. 选择创建启动模板。
16. 在确认页上，选择 Create Auto Scaling group (创建 Auto Scaling 组)。

如果您当前未使用启动模板，您可以转而创建启动配置。

启动配置类似于启动模板，它指定 Amazon EC2 Auto Scaling 为您创建的 EC2 实例的类型。请包含一些信息以创建启动配置，例如，要使用的 Amazon 系统映像 (AMI) 的 ID、实例类型、密钥对、安全组和块储存设备映射。

创建启动配置

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航栏上，选择一个区域。您创建的 Auto Scaling 资源与您所指定的区域相关联。
3. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling Groups (Auto Scaling 组)。
4. 在 Welcome to Auto Scaling (欢迎使用 Auto Scaling) 页面上，选择 Create Auto Scaling group (创建 Auto Scaling 组)。



5. 在 Create Auto Scaling Group (创建 Auto Scaling 组) 页面上，选择 Launch Configuration, Create a new launch configuration (启动配置、创建新启动配置)，然后选择 Next Step (下一步)。
6. 对于 Choose AMI (选择 AMI) 步骤，有一组称为 Amazon 系统映像 (AMI) 的基本配置，它们充当您的实例的模板。对于 Amazon Linux 2 AMI，选择 Select (选择)。
7. 对于 Choose Instance Type (选择实例类型) 步骤，为您的实例选择硬件配置。建议您保留默认值，即 t2.micro 实例。选择 Next: Configure details (下一步：配置详细信息)。
8. 对于 Configure details (配置详细信息) 步骤，执行以下操作：
 - a. 对于 Name，为启动配置键入一个名称（例如，my-first-lc）。
 - b. 对于 Advanced Details，选择 IP 地址类型。如果为 VPC 中的实例提供 Internet 连接，您必须选择一个指定了公有 IP 地址的选项。如果您要向实例提供 Internet 连接，但是不确定您是否有默认 VPC，请选择向每个实例分配公有 IP 地址。
 - c. 选择 Skip to review。
9. 对于 Review (审核) 步骤，选择 Edit security groups (编辑安全组)。按照说明操作来选择现有安全组，然后选择 Review。
10. 对于 Review (审核) 步骤，选择 Create launch configuration (创建启动配置)。
11. 按照提示完成 Select an existing key pair or create a new key pair (选择现有密钥对或创建新密钥对)。在本教程中，您不会连接到实例。因此，您可以选择 Proceed without a key pair，除非您打算连接到您的实例。
12. 选择 Create launch configuration。创建启动配置并显示创建 Auto Scaling 组的向导。

步骤 2：创建 Auto Scaling 组

Auto Scaling 组是一个 EC2 实例集合，并且是 Amazon EC2 Auto Scaling 的核心。在创建 Auto Scaling 组时，您需要包含一些信息，例如，实例的子网和开始时的初始实例数。

创建启动配置或模板后，使用以下过程从中断的地方继续操作。

创建 Auto Scaling 组

1. 对于 Configure Auto Scaling group details (配置 Auto Scaling 组详细信息) 步骤，执行以下操作：
 - a. 对于 Group name (组名)，键入您的 Auto Scaling 组的名称（例如，my-first-asg）。

- [启动模板] 对于启动模板版本，选择 Auto Scaling 组在扩展时使用启动模板的默认版本、最新版本还是特定版本。
- 对于队列构成，选择遵循启动模板。
- 在本教程中，将 Group size 保留为默认值，即 1 个实例。
- 将网络设置为您选择的 AWS 区域的默认 VPC，或者选择您自己的 VPC。
- 对于子网，选择 VPC 的一个子网。

Note

您可以通过选择相应的默认子网来选择实例的可用区。

- 选择 Next: Configure scaling policies。
- 在 Configure scaling policies 页面上，选择 Keep this group at its initial size 和 Review。
 - 在 Review 页面上，选择 Create Auto Scaling group。
 - 在 Auto Scaling group creation status 页面上，选择 Close。

步骤 3：验证 Auto Scaling 组

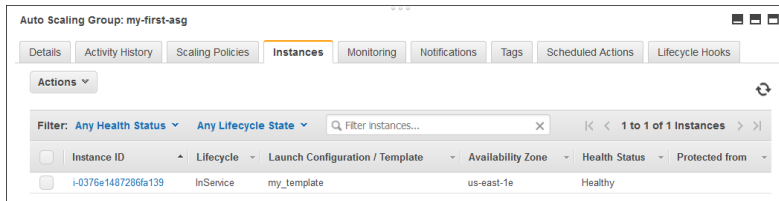
现在，您已创建您的 Auto Scaling 组，可验证该组是否已启动 EC2 实例。

验证您的 Auto Scaling 组是否已启动 EC2 实例

- 在 Auto Scaling Groups 页面上，选择您刚创建的 Auto Scaling 组。
- Details 选项卡提供了有关 Auto Scaling 组的信息。

- 在 Activity History 选项卡上，Status 列显示实例的当前状态。当您的实例启动时，状态列将显示 In progress。该实例启动后，状态会变为 Successful。您还可以使用刷新按钮来查看您的实例的当前状态。

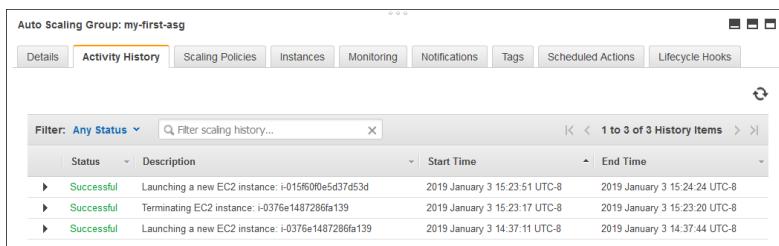
- 在 Instances 选项卡上，Lifecycle 列显示实例的状态。您可以看到 Auto Scaling 组已启动 EC2 实例，并且该实例处于 InService 生命周期状态。Health Status (运行状态) 列显示对您的实例进行 EC2 实例运行状况检查的结果。



(可选) 终止您的 Auto Scaling 组中的实例

如果需要，您可以尝试以下实验以了解 Amazon EC2 Auto Scaling 的更多信息。Auto Scaling 组的最小大小为 1 个实例。因此，如果您终止了正在运行的实例，Amazon EC2 Auto Scaling 必须启动一个新实例替换它。

- 在 Instances 选项卡上，选择实例的 ID。这将在 Instances 页面上显示实例。
- 依次选择 Actions、Instance State 和 Terminate。当系统提示您确认时，选择 Yes, Terminate。
- 在导航窗格中，选择 Auto Scaling Groups (Auto Scaling 组)。选择您的 Auto Scaling 组，然后选择 Activity History (活动历史记录) 选项卡。Auto Scaling 组的默认冷却时间是 300 秒 (5 分钟)，因此您大约需要等 5 分钟才能看到扩展活动。在扩展活动启动时，将会看到有关终止第一个实例的条目以及有关启动新实例的条目。Instances 选项卡仅显示新实例。



- 在导航窗格中，选择实例。此页面同时显示已终止的实例和正在运行的实例。

如果要删除自动扩展的基本基础设施，请转到下一步。否则，您可以将该基础设施作为基础，然后尝试以下一个或多个操作：

- [Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)
- [适用于 Amazon EC2 Auto Scaling 的动态扩展 \(p. 76\)](#)
- [在 Auto Scaling 组扩展时获取 Amazon SNS 通知 \(p. 127\)](#)

步骤 4：(可选) 删除扩展基础设施

您可以删除扩展基础设施，或者仅删除 Auto Scaling 组而保留启动模板以供将来使用。

如果您启动的实例不在 [AWS 免费套餐](#) 范围内，则应终止实例以避免产生额外费用。将删除 EC2 实例以及相关数据。

删除 Auto Scaling 组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling Groups (Auto Scaling 组)。

3. 选择您的 Auto Scaling 组。
4. 依次选择 Actions 和 Delete。当系统提示进行确认时，选择 Yes, Delete。

Name 列指示 Auto Scaling 组正在被删除。所需、最小和最大列显示 Auto Scaling 组具有 0 个实例。

如果要保留启动模板，请跳过该过程。

删除启动模板

1. 在导航窗格中，选择启动模板。
2. 选择启动模板。
3. 选择操作，然后选择删除模板。在提示确认时，选择删除启动模板。

如果您要保留启动配置，请跳过这一步。

删除启动配置

1. 在导航窗格上的 Auto Scaling 下，选择 Launch Configurations (启动配置)。
2. 选择启动配置。
3. 依次选择 Actions 和 Delete launch configuration。当系统提示进行确认时，选择 Yes, Delete。

教程：设置具有扩展和负载均衡功能的应用程序

可以将负载均衡器附加到 Auto Scaling 组。该负载均衡器在组中的所有实例间自动分配传入流量。

在本教程中，在创建 Auto Scaling 组时，可以向该组附加负载均衡器。要将负载均衡器附加到现有 Auto Scaling 组，请参阅[将负载均衡器附加到 Auto Scaling 组 \(p. 54\)](#)。

在开始学习本教程之前，我们建议您首先阅读以下介绍性主题：[将负载均衡器与 Auto Scaling 组结合使用 \(p. 53\)](#)。

目录

- [先决条件 \(p. 17\)](#)
- [配置扩展和负载均衡功能 \(控制台\) \(p. 17\)](#)
- [配置扩展和负载均衡功能 \(AWS CLI\) \(p. 20\)](#)
- [清除您的 AWS 资源 \(p. 21\)](#)

先决条件

- (可选) 创建一个 IAM 角色，该角色将为您的应用程序授予对所需 AWS 的访问权限。
- 启动一个实例。务必指定 IAM 角色 (如果已创建) 并将所需的任何配置脚本指定为用户数据。连接到实例并对其进行自定义。例如，您可以安装软件 and 应用程序并复制数据。测试您的实例上的应用程序以确保您的实例配置正确。从您的实例创建自定义 Amazon 系统映像 (AMI)。如果您不再需要实例，可终止它。
- 创建负载均衡器。Elastic Load Balancing 支持三种类型的负载均衡器：Application Load Balancer、Network Load Balancer 和 Classic Load Balancer。您可以将任一类型的负载均衡器附加到 Auto Scaling 组。有关更多信息，请参阅 [Elastic Load Balancing 类型 \(p. 53\)](#)。
- 确保您针对计划为 Auto Scaling 组启用的负载均衡器选择了相同的可用区。

配置扩展和负载均衡功能 (控制台)

在创建您的 Auto Scaling 组时，请完成以下任务以设置具有扩展和负载均衡功能的应用程序。

任务

- [创建或选择启动配置 \(p. 17\)](#)
- [创建或选择启动模板 \(p. 18\)](#)
- [创建 Auto Scaling 组 \(p. 19\)](#)
- (可选) [验证您的负载均衡器是否已附加到您的 Auto Scaling 组 \(p. 20\)](#)

创建或选择启动配置

启动配置会指定 Amazon EC2 Auto Scaling 为您创建的 EC2 实例类型。在创建启动配置时，您需要包含要使用的 Amazon 系统映像 (AMI) 的 ID、实例类型、密钥对和块储存设备映射等信息。如果创建了启动模板，

您可以使用启动模板创建 Auto Scaling 组而不是使用启动配置。有关更多信息，请参阅 [创建或选择启动模板 \(p. 18\)](#)。

如果您已拥有要使用的启动配置，请使用以下过程选择该启动配置。

选择现有的启动配置

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏上，选择在创建负载均衡器时使用的 AWS 区域。
3. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
4. 在下一页上，选择 Create Auto Scaling group (创建 Auto Scaling 组)。
5. 在 Create Auto Scaling Group (创建 Auto Scaling 组) 页面上，选择 Launch Configuration (启动配置)，选择现有启动配置，然后选择 Next Step (下一步)。

要创建新的启动配置，请使用下列过程：

创建启动配置

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏上，选择在创建负载均衡器时使用的 AWS 区域。
3. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
4. 在下一页上，选择 Create Auto Scaling group (创建 Auto Scaling 组)。
5. 在创建 Auto Scaling 组页面上，选择启动配置、创建新启动配置，然后选择下一步。
6. 在 Choose AMI 页面上，选择您的自定义 AMI。
7. 在 Choose Instance Type 页面上，选择实例的硬件配置，然后选择 Next: Configure details。
8. 在 Configure Details (配置实例详细信息) 页面中，执行以下操作：
 - a. 对于名称，为您的启动配置输入名称。
 - b. (可选) 要安全地将凭证分配给您的 EC2 实例，请选择您的 IAM 角色。
 - c. (可选) 如果需要连接到非默认 VPC 内的实例，则对于高级详细信息和 IP 地址类型，选择向每个实例分配公有 IP 地址。
 - d. (可选) 要为实例指定用户数据或配置脚本，则对于 Advanced Details 和 User data，粘贴您的配置脚本。
 - e. 选择 Skip to review。
9. 在 Review 页面上，选择 Edit security groups。按照说明操作来选择现有安全组，然后选择 Review。
10. 在 Review 页上选择 Create launch configuration。
11. 在 Select an existing key pair or create a new key pair 页面上，选择所列的选项之一。选中确认复选框，然后选择 Create launch configuration。

Warning

如果需要连接到您的实例，不要选择 Proceed without a key pair。

按照上述说明完成操作之后，您已准备好继续使用向导来创建 Auto Scaling 组。

创建或选择启动模板

如果您已拥有要使用的启动模板，请使用以下过程选择该启动模板。

选择现有启动模板

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。

2. 在屏幕顶部的导航栏上，选择在创建负载均衡器时使用的 AWS 区域。
3. 在导航窗格中，选择启动模板。
4. 选择启动模板。
5. 选择 Actions (操作)、Create Auto Scaling group (创建 Auto Scaling 组)。

或者，使用下列过程创建新的启动模板。

创建启动模板

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏上，选择在创建负载均衡器时使用的 AWS 区域。
3. 在导航窗格中，选择启动模板。
4. 选择创建启动模板。
5. 提供启动模板的名称和描述。启动模板名称必须包含 3 到 125 个字符，可以包含以下字符：()/_。
6. 对于 AMI ID，输入您的实例的 AMI ID。
7. 对于实例类型，选择与您指定的 AMI 兼容的实例硬件配置。
8. (可选) 对于密钥对名称，键入在连接到您的实例时使用的密钥对的名称。
9. 对于网络接口，执行以下操作：
 - a. 选择添加网络接口。
 - b. (可选) 要连接到非默认 VPC 中的实例，对于自动分配公有 IP，选择启用。
 - c. 对于安全组 ID，为实例指定安全组。
 - d. 对于终止时删除，请选择在 Auto Scaling 组缩减并终止网络接口附加到的实例时是否删除网络接口。
10. (可选) 要将凭证安全地分配到实例，对于高级详细信息和 IAM 实例配置文件，输入 IAM 角色的 Amazon 资源名称 (ARN)。
11. (可选) 要为实例指定用户数据或配置脚本，请将其粘贴到高级详细信息和用户数据。
12. 选择创建启动模板。
13. 在确认页上，选择 Create Auto Scaling group (创建 Auto Scaling 组)。

创建 Auto Scaling 组

选择或创建启动配置或模板后，使用以下步骤从中断的地方继续操作。

创建 Auto Scaling 组

1. 在 Configure Auto Scaling group details 页面上，执行以下操作：
 - a. 对于组名，输入 Auto Scaling 组的名称。
 - b. [启动模板] 对于启动模板版本，选择 Auto Scaling 组在扩展时使用启动模板的默认版本、最新版本还是特定版本。
 - c. [启动模板] 对于队列构成，请选择遵循启动模板以使用在启动模板中指定的 EC2 实例类型和购买选项。
 - d. 对于组大小，输入 Auto Scaling 组的初始实例数。
 - e. 如果为启动配置或模板选择了需要 VPC 的实例类型（例如 T2 实例），则必须为网络选择 VPC。否则，如果您的账户支持 EC2-Classic 而且您已选择不需要 VPC 的实例类型，则可以选择 Launch into EC2-Classic 或 VPC。
 - f. 如果您在上一步中选择了 VPC，请从 Subnet 中选择一个或多个子网。如果您选择 EC2-Classic，请从可用区中选择一个或多个可用区。

- g. 对于高级详细信息，选择从弹性负载均衡器接收流量，然后执行以下操作之一：
 - [Classic 负载均衡器] 从 Load Balancers 中选择您的负载均衡器。
 - [应用程序/网络负载均衡器] 从目标组中选择您的目标组。
 - h. (可选) 要使用 Elastic Load Balancing 运行状况检查，请为 Advanced Details (高级详细信息) 和 Health Check Type (运行状况检查类型) 选择 ELB。
 - i. 选择 Next: Configure scaling policies。
2. 在 Configure scaling policies 页面上，选择 Keep this group at its initial size，然后选择 Review。

要为您的 Auto Scaling 组配置扩展策略，请参阅[配置扩展策略 \(p. 79\)](#)。
 3. 审核 Auto Scaling 组的详细信息。您可以选择 Edit 进行更改。完成后，选择 Create Auto Scaling group (创建 Auto Scaling 组)。

(可选) 验证您的负载均衡器是否已附加到您的 Auto Scaling 组

验证您的负载均衡器是否已附加到您的 Auto Scaling 组

1. 选择您的 Auto Scaling 组。
2. 在 Details 选项卡上，Load Balancers 显示任何附加的负载均衡器，Target Groups 显示任何附加的目标组。
3. 在 Activity History (活动历史记录) 选项卡上，Status (状态) 列显示 Auto Scaling 实例的状态。实例启动时，其状态是 In progress。该实例启动后，状态会变为 Successful。
4. 在 Instances (实例) 选项卡上，Lifecycle (生命周期) 列显示 Auto Scaling 实例的状态。在实例准备好接收流量时，其状态为 InService。

Health Status 列显示了对您的实例执行运行状况检查的结果。

配置扩展和负载均衡功能 (AWS CLI)

完成以下任务以设置具有扩展和负载均衡功能的应用程序。

任务

- [创建启动配置 \(p. 20\)](#)
- [创建启动模板 \(p. 21\)](#)
- [创建包含负载均衡器的 Auto Scaling 组 \(p. 21\)](#)

创建启动配置

如果您已经有了要使用的启动配置，请跳过这一步。

创建启动配置的步骤

使用以下 [create-launch-configuration](#) 命令。要为非默认 VPC 中的实例分配公有 IP 地址，请如下所示使用 --associate-public-ip-address 选项。

```
aws autoscaling create-launch-configuration --launch-configuration-name my-launch-config \
--image-id ami-01e24be29428c15b2 --instance-type t2.micro --associate-public-ip-address \
--security-groups sg-eb2af88e
```

创建启动模板

如果您已经有了要使用的启动模板，请跳过这一步。

创建启动模板

使用以下 `create-launch-template` 命令，并传递一个包含用于创建启动模板的信息的 JSON 文件。要为非默认 VPC 中的实例分配公有 IP 地址，请添加网络接口属性并将其设置为 `"NetworkInterfaces": {"AssociatePublicIpAddress": true}`，如下所示。

```
aws ec2 create-launch-template --launch-template-name my-launch-template --version-  
description my-version-description \  
  --launch-template-data '{"NetworkInterfaces":  
[{"DeviceIndex":0,"AssociatePublicIpAddress":true,"Groups":  
["sg-903004f8"],"DeleteOnTermination":true}], "ImageId":"ami-01e24be29428c15b2", "InstanceType":"t2.micro"
```

创建包含负载均衡器的 Auto Scaling 组

在创建 Auto Scaling 组时，您可以为该组挂载一个现有的负载均衡器。您可以使用启动配置或启动模板来自自动配置 Auto Scaling 组启动的实例。

创建带附加 传统负载均衡器 的 Auto Scaling 组

可以将以下 `create-auto-scaling-group` 命令与 `--load-balancer-names` 选项一起使用以创建一个附加了传统负载均衡器的 Auto Scaling 组。

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
  --launch-configuration-name my-launch-config \  
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  
  --load-balancer-names "my-load-balancer" \  
  --max-size 5 --min-size 1 --desired-capacity 2
```

为 应用程序负载均衡器 或 网络负载均衡器 创建带附加的目标组的 Auto Scaling 组

可以将以下 `create-auto-scaling-group` 命令与 `--target-group-arns` 选项一起使用以创建一个附加了目标组的 Auto Scaling 组。

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
  --launch-template "LaunchTemplateName=my-launch-template,Version=1" \  
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  
  --target-group-arns "arn:aws:elasticloadbalancing:region:123456789012:targetgroup/my-  
targets/1234567890123456" \  
  --max-size 5 --min-size 1 --desired-capacity 2
```

清除您的 AWS 资源

现在您已成功完成了教程。为了避免未使用的资源持续产生费用，您应清除刚刚为此教程创建的资源。如需分步指导，请参阅 [删除 Auto Scaling 基础设施](#) (p. 61)。

启动模板

启动模板类似于[启动配置](#) (p. 28)，因为它指定实例配置信息。包括 Amazon 系统映像 (AMI) ID、实例类型、密钥对、安全组以及用于启动 EC2 实例的其他参数。但是，定义启动模板而非启动配置可让您有多个版本的模板。利用版本控制，您可以创建完整一组参数的一个子集，然后重复使用它创建其他模板或模板版本。例如，您可以创建一个默认模板来定义常用的配置参数（如标签或网络配置），并允许将其他参数指定为同一模板另一个版本的一部分。

我们建议您使用启动模板而不是启动配置，以确保您可以使用最新的 Amazon EC2 功能，例如 T2 Unlimited 实例。有关更多信息，请参阅[Amazon EC2 用户指南](#)（适用于 Linux 实例）中的[突发性能实例的无限模式](#)和[使用 Auto Scaling 组在无限模式下启动突发性能实例](#)。

使用启动模板，您还可以同时使用按需实例与 Spot 实例跨多个实例类型预置容量，以实现所需的规模、性能和成本。有关更多信息，请参阅[Auto Scaling 组具有多个实例类型和购买选项](#) (p. 37)。

如果当前使用启动配置，您可以在更新使用启动配置创建的 Auto Scaling 组时指定启动模板。

要创建用于 Auto Scaling 组的启动模板，请从头开始创建模板，创建现有模板的新版本，或者从启动配置、运行的实例或其他模板中复制参数。

以下主题介绍用于创建和使用用于 Auto Scaling 组的启动模板的最常见过程。有关启动模板的更多信息，请参阅[Amazon EC2 用户指南](#)（适用于 Linux 实例）中的[启动模板](#)部分。

目录

- [为 Auto Scaling 组创建启动模板](#) (p. 22)
- [将启动配置复制到启动模板](#) (p. 26)
- [使用启动模板替换启动配置](#) (p. 27)

为 Auto Scaling 组创建启动模板

在可以使用启动模板创建 Auto Scaling 组之前，您必须创建启动模板，其中包括启动 EC2 实例所需的参数，例如 Amazon 系统映像 (AMI) 的 ID 和实例类型。

以下过程用于创建新的启动模板。新模板使用您定义的参数（从头开始）或现有的启动模板。在创建启动模板之前，您可以按照[使用启动模板创建 Auto Scaling 组](#) (p. 44)中的说明创建 Auto Scaling 组。

先决条件

有关所需的 IAM 权限的信息，请参阅[Amazon EC2 用户指南](#)（适用于 Linux 实例）中的[控制如何使用启动模板](#)。

注意事项

在创建用于 Auto Scaling 组的启动模板时，请记住以下注意事项：

- 启动模板为您提供了灵活性，让您选择是启动一种类型的实例，还是将启动实例类型与按需和 Spot 购买选项组合。有关更多信息，请参阅[Auto Scaling 组具有多个实例类型和购买选项](#) (p. 37)。不支持启动具有以下组合的实例：
 - 如果在其他详细信息中指定 Spot 实例请求
 - 在 EC2-Classic 中
- 如果您为模板配置网络类型（VPC 或 EC2-Classic）、子网和可用区，将忽略这些设置以支持在 Auto Scaling 组中指定的设置。
- 如果您指定一个网络接口，则必须将安全组配置为该网络接口的一部分，而不是在模板的 Security Groups 部分中指定。
- 您不能指定多个网络接口。

- 您不能分配特定的私有 IP 地址。当实例启动时，会在从其中启动实例的子网的 CIDR 范围中分配私有地址。有关为您的 VPC 或子网指定 CIDR 范围的更多信息，请参阅[Amazon VPC 用户指南](#)。
- 要指定要使用的现有网络接口，其设备索引必须为 0 (eth0)。在这种情况下，您必须使用 CLI 或 API 来创建 Auto Scaling 组。当您使用 CLI create-auto-scaling-group 命令或 API CreateAutoScalingGroup 操作创建组时，您必须指定可用区参数而不是子网（VPC 区标识符）参数。
- 您不能使用主机放置关联。

使用控制台为 Auto Scaling 组创建新的启动模板

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Launch Templates。
3. 选择创建新模板。为启动模板的初始版本输入名称并提供描述。
4. 如果您选择基于另一个模板创建新的模板：
 - a. 对于源模板，请选择现有启动模板。
 - b. 从源模板版本中选择新启动模板版本所基于的启动模板版本。
5. 在启动模板内容下，请提供以下信息。
 - a. AMI ID：选择要在其上启动实例的 Amazon 系统映像 (AMI) 的 ID。您可以使用 Search for AMI (搜索 AMI) 对话框搜索所有可用 AMI。从 Quick Start (快速启动) 列表选择一个常用的 AMI。如果您没有看到所需的 AMI，请选择 AWS Marketplace 或 Community AMIs (社区 AMI) 列表来[查找合适的 AMI](#)。
 - b. 实例类型：选择[实例类型](#)。确保实例类型与指定的 AMI 兼容。
 - c. 密钥对名称：指定连接到实例时要使用的[密钥对](#)。
 - d. 网络类型：您可以选择指定是将实例启动到 VPC 还是 EC2-Classic (如果适用)。不过，Amazon EC2 Auto Scaling 将忽略启动模板的网络类型和可用区设置以支持 Auto Scaling 组设置。
 - e. 安全组：选择一个或多个[安全组](#)，或留空以将安全组配置为网络接口的一部分。您不能在这两处指定安全组。如果要使用 EC2-Classic，则必须使用专为 EC2-Classic 创建的安全组。
6. 在网络接口下面，选择添加网络接口并提供以下可选的信息。您只能指定一个网络接口。请注意以下字段：
 - a. 设备：指定 eth0 作为设备名称（所用设备的设备索引为 0）。
 - b. 网络接口：留空以让 AWS 在启动实例时创建新的网络接口，或者输入现有网络接口的 ID。如果您指定了 ID，则这会将您的 Auto Scaling 组限制为一个实例。
 - c. 描述：输入一个描述性名称。
 - d. Subnet (子网)：虽然您可以选择指定一个子网，但 Amazon EC2 Auto Scaling 会忽略该子网以支持 Auto Scaling 组设置。
 - e. 自动分配公有 IP：指定是否自动为设备索引为 eth0 的网络接口分配公有 IP 地址。只能为单个新网络接口启用该设置。
 - f. Security group ID (安全组 ID)：输入要与主网络接口 (eth0) 关联的一个或多个[安全组](#)的 ID。对于 Auto Scaling 组会将实例启动到其中的 VPC，必须为其配置各个安全组。条目之间用逗号分隔。
 - g. 终止时删除：选择在 Auto Scaling 组缩减并终止网络接口附加到的实例时是否删除网络接口。
7. 如果您选择指定要附加到实例的卷，而不包含 AMI 指定的卷，则在存储（卷）下提供以下信息：
 - a. 卷类型：选择实例存储或 Amazon EBS 卷。卷类型取决于您选择的实例类型。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[Amazon EC2 实例存储](#)和[Amazon EBS 卷](#)。
 - b. 设备名称：指定卷的设备名称。
 - c. 快照：输入用于创建卷的快照的 ID。
 - d. 大小：对于 Amazon EBS 支持的卷，请指定存储大小。如果您要从快照中创建卷，但未指定卷大小，则默认值为快照大小。
 - e. 卷类型：对于 Amazon EBS 卷，请选择[卷类型](#)。

- f. IOPS：对于预置的 IOPS SSD 卷，输入该卷应该支持的每秒输入/输出操作 (IOPS) 的最大数量。
- g. 终止时删除：对于 Amazon EBS 卷，选择在终止关联实例时是否删除此卷。
- h. 加密：选择是可更改 Amazon EBS 卷的加密状态。设置此参数的默认效果会随所选卷源而异，如下表所述。在所有情况下，您都必须拥有使用指定 CMK 的权限。有关指定加密卷的更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [Amazon EBS 加密](#)。

加密结果

如果 Encrypted 参数设置为...	如果卷源...	则默认加密状态为...	备注
否	新（空）卷	未加密*	不适用
	您拥有的未加密快照	未加密*	
	您拥有的加密快照	使用相同密钥加密	
	与您共享的未加密快照	未加密*	
	与您共享的加密快照	加密到默认 CMK	
是	新卷	加密到默认 CMK	要使用非默认 CMK，请为密钥参数指定一个值。
	您拥有的未加密快照	加密到默认 CMK	
	您拥有的加密快照	使用相同密钥加密	
	与您共享的未加密快照	加密到默认 CMK	
	与您共享的加密快照	加密到默认 CMK	

* 如果启用了 [默认加密](#)，则所有新创建的卷（不论是否将加密参数设置为是）将使用默认 CMK 加密。同时设置加密和密钥参数可指定非默认 CMK。

- i. [可选] 密钥：如果您在上一步中选择了是，则输入在对卷加密时要使用的客户主密钥 (CMK)。您可以输入以前使用 AWS Key Management Service 创建的任何 CMK。您可以粘贴您可以访问的任何密钥的完整 ARN。有关更多信息，请参阅 [AWS Key Management Service Developer Guide](#) 和本指南中的 [与加密卷结合使用时必需的 CMK 密钥策略](#) (p. 145)。

Note

提供 CMK 但未同时设置加密参数会导致错误。

- 8. 对于实例标签，请提供键和值组合以指定标签。您可以标记实例、卷或两者。
- 9. 对于高级详细信息，请展开该部分以查看字段并为实例指定任何其他参数。有关本节以及如何使用每个参数的更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [创建启动模板](#)。
- 购买选项：您可以选择请求 Spot 实例，并指定您愿意支付的最高实例每小时价格。要使用 Auto Scaling 组来完成此事项，您必须指定不带结束日期的一次性请求。有关更多信息，请参阅 [在 Auto Scaling 组中启动 Spot 实例](#) (p. 59)。

Important

如果您计划在配置 Auto Scaling 组时指定多个实例类型和购买选项，请将这些字段留空。有关更多信息，请参阅 [Auto Scaling 组具有多个实例类型和购买选项](#) (p. 37)。

- IAM 实例配置文件：指定要与实例关联的 AWS Identity and Access Management (IAM) 实例配置文件。有关更多信息，请参阅 [在 Amazon EC2 实例上运行的应用程序的 IAM 角色](#) (p. 147)。
- 关闭操作：您可以将此字段留空，因为 Amazon EC2 Auto Scaling 将忽略它。Amazon EC2 Auto Scaling 的默认行为是终止实例。

- 终止保护：提供额外的终止保护，但是在扩展 Auto Scaling 组时，Amazon EC2 Auto Scaling 会将其忽略。要控制 Auto Scaling 组在扩展时是否可以终止特定实例，请使用[实例保护](#) (p. 98)。
- 监控：选择是否使用 Amazon CloudWatch 启用实例的详细监控。将收取额外费用。有关更多信息，请参阅[使用 Amazon CloudWatch 监控 Auto Scaling 组和实例](#) (p. 117)。
- T2/T3 无限：（仅对 T2 和 T3 实例有效）选择是否允许应用程序突增到基准以上并持续所需的时间。可能收取额外费用。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[使用 Auto Scaling 组以无限模式启动突发性能实例](#)。
- 置放群组名称：指定在其中启动实例的置放群组。并非可以在置放群组中启动所有实例类型。如果使用指定其他置放群组的 CLI 命令配置 Auto Scaling 组，将忽略该设置以支持为 Auto Scaling 组指定的设置。
- EBS 优化的实例：为 Amazon EBS I/O 提供额外的专用容量。并非所有实例类型都支持该功能，并且会产生额外的费用。
- 租期：您可以选择是在共享硬件（共享），还是在隔离的专用硬件（专用）上运行您的实例。可能收取额外费用。您无法选择专用主机。
- RAM 磁盘 ID：与 AMI 关联的 RAM 磁盘的 ID。仅对半虚拟化 (PV) AMI 有效。
- 内核 ID：与 AMI 关联的内核的 ID。仅对半虚拟化 (PV) AMI 有效。
- User data：您可以指定用户数据在启动时配置实例或运行配置脚本。

10. 选择创建启动模板。

通过现有实例创建启动模板

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择实例。
3. 选择实例，然后选择操作、Create template from instance (从实例创建模板)。
4. 提供名称和说明。根据需要，调整任何其他启动参数，然后选择创建启动模板。

使用命令行创建启动模板

您可以使用以下任一命令：

- [create-launch-template](#) (AWS CLI)
- [New-EC2LaunchTemplate](#) (适用于 Windows PowerShell 的 AWS 工具)

使用 [create-launch-template](#) 命令行创建启动模板，如下所示。为 Groups 指定一个值，它与 Auto Scaling 组将实例启动到其中的 VPC 的安全组对应。指定 VPC 和子网作为 Auto Scaling 组的属性。

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
--launch-template-data '{"NetworkInterfaces":
[{"DeviceIndex":0,"AssociatePublicIpAddress":true,"Groups":
["sg-7c227019"],"DeleteOnTermination":true},"ImageId":"ami-01e24be29428c15b2","InstanceType":"t2.micro",
[{"ResourceType":"instance","Tags":[{"Key":"purpose","Value":"webserver"}]}]}'
```

使用以下 [describe-launch-templates](#) 命令描述启动模板 [my-template-for-auto-scaling](#)。

```
aws ec2 describe-launch-templates --launch-template-names my-template-for-auto-scaling
```

使用以下 [describe-launch-template-versions](#) 命令描述指定启动模板 [my-template-for-auto-scaling](#) 的版本。

```
aws ec2 describe-launch-template-versions --launch-template-id lt-068f72b72934aff71
```

以下为响应示例：

```
{
  "LaunchTemplateVersions": [
    {
      "VersionDescription": "version1",
      "LaunchTemplateId": "lt-068f72b72934aff71",
      "LaunchTemplateName": "my-template-for-auto-scaling",
      "VersionNumber": 1,
      "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
      "LaunchTemplateData": {
        "TagSpecifications": [
          {
            "ResourceType": "instance",
            "Tags": [
              {
                "Key": "purpose",
                "Value": "webserver"
              }
            ]
          }
        ],
        "ImageId": "ami-01e24be29428c15b2",
        "InstanceType": "t2.micro",
        "NetworkInterfaces": [
          {
            "DeviceIndex": 0,
            "DeleteOnTermination": true,
            "Groups": [
              "sg-7c227019"
            ],
            "AssociatePublicIpAddress": true
          }
        ]
      },
      "DefaultVersion": true,
      "CreateTime": "2019-02-28T19:52:27.000Z"
    }
  ]
}
```

将启动配置复制到启动模板

使用以下过程，从现有启动配置复制选项以创建新启动模板。

您可以从现有启动配置创建启动模板，以便您轻松更新 Auto Scaling 组来使用启动模板。与启动配置类似，启动模板可以包含用于启动实例的全部或部分参数。使用启动模板，您还可以创建模板的多个版本，以便更快地启动新实例。

从启动配置创建启动模板

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Launch Configurations。
3. 选择您要复制的启动配置，然后选择复制到启动模板。这将设置一个与您所选启动配置具有相同名称和选项的新启动模板。
4. 对于新建启动模板名称，您可以使用启动配置的名称（默认值）或者输入新名称。启动模板名称必须是唯一的。
5. （可选）若要使用新启动模板创建 Auto Scaling 组，请选择使用新模板创建 Auto Scaling 组。有关更多信息，请参阅 [使用启动模板创建 Auto Scaling 组 \(p. 44\)](#)。

6. 选择 Submit。

创建启动模板之后，您可以使用所创建的启动模板，根据需要更新现有 Auto Scaling 组。有关更多信息，请参阅 [使用启动模板替换启动配置 \(p. 27\)](#)。

使用启动模板替换启动配置

编辑具有现有启动配置的 Auto Scaling 组时，您可以选择使用启动模板替换启动配置。这使您可以将启动模板用于当前使用的任意 Auto Scaling 组。在这样做的过程中，您可以利用启动模板的版本控制和其他功能。

替换 Auto Scaling 组的启动配置后，启动的任何新实例将使用新的启动模板，但是现有实例不受影响。

先决条件

您必须先创建启动模板，然后才能替换 Auto Scaling 组中的启动配置。创建启动模板最简单的方法从启动配置进行复制。有关更多信息，请参阅 [将启动配置复制到启动模板 \(p. 26\)](#)。

替换 Auto Scaling 组的启动配置

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Auto Scaling Groups。
3. 选择 Auto Scaling 组，然后选择详细信息、编辑。
4. 对于使用以下方式启动实例，选择启动模板选项。
5. 对于启动模板，选择您的启动模板。
6. 对于启动模板版本，根据需要选择启动模板版本。在创建启动模板版本之后，您可以选择 Auto Scaling 组在扩展时是使用启动模板的默认版本还是最新版本。
7. 完成后，选择 Save。

使用命令行替换启动配置

您可以使用以下任一命令：

- [update-auto-scaling-group](#) (AWS CLI)
- [Update-ASAutoScalingGroup](#) (适用于 Windows PowerShell 的 AWS 工具)

启动配置

启动配置是 Auto Scaling 组用于启动 EC2 实例的实例配置模板。在创建启动配置时，您需要指定实例的信息。包括 Amazon 系统映像 (AMI) ID、实例类型、密钥对、一个或多个安全组以及块储存设备映射。如果之前已启动过 EC2 实例，可以指定相同的信息来启动实例。

您可以为多个 Auto Scaling 组指定启动配置。但是一次只能为一个 Auto Scaling 组指定一个启动配置，而且启动配置在创建后不能修改。要更改 Auto Scaling 组的启动配置，您必须创建一个启动配置，然后使用该配置更新 Auto Scaling 组。

请记住，只要在您创建或更新 Auto Scaling 组时，就必须指定启动配置、启动模板或 EC2 实例。当您使用 EC2 实例创建 Auto Scaling 组时，Amazon EC2 Auto Scaling 自动创建启动配置并将其与 Auto Scaling 组关联。有关更多信息，请参阅[使用 EC2 实例创建 Auto Scaling 组 \(p. 46\)](#)。或者，如果您使用启动模板，您可以指定启动模板而非启动配置或 EC2 实例。有关更多信息，请参阅[启动模板 \(p. 22\)](#)。

目录

- [创建启动配置 \(p. 28\)](#)
- [使用 EC2 实例创建启动配置 \(p. 29\)](#)
- [更改 Auto Scaling 组的启动配置 \(p. 33\)](#)
- [在 VPC 中启动 Auto Scaling 实例 \(p. 33\)](#)

创建启动配置

在创建启动配置时，您必须指定有关要启动的 EC2 实例的信息。包括 Amazon 系统映像 (AMI)、实例类型、密钥对、安全组和块储存设备映射。或者，您可以使用正在运行的 EC2 实例的属性来创建启动配置。有关更多信息，请参阅[使用 EC2 实例创建启动配置 \(p. 29\)](#)。

Note

在创建 Auto Scaling 组时，您可以指定启动模板、启动配置或 EC2 实例。建议您使用启动模板以确保您可以使用 Amazon EC2 的最新功能。有关更多信息，请参阅[启动模板 \(p. 22\)](#)。

在创建启动配置后，您可创建 Auto Scaling 组。有关更多信息，请参阅[使用启动配置创建 Auto Scaling 组 \(p. 45\)](#)。

Auto Scaling 组一次与一个启动配置关联，并且启动配置在创建后不能修改。因此，如果您要更改现有 Auto Scaling 组的启动配置，则必须使用新的启动配置更新它。有关更多信息，请参阅[更改 Auto Scaling 组的启动配置 \(p. 33\)](#)。

使用控制台创建启动配置

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏上，将显示当前区域。为 Auto Scaling 组选择一个符合您需求的区域。
3. 在导航窗格上的 Auto Scaling 下，选择 Launch Configurations (启动配置)。
4. 在下一页上，选择创建启动配置。
5. 在 Choose AMI 页面中，选择 AMI。
6. 在 Choose Instance Type 页面上，为您的实例选择硬件配置。选择 Next: Configure details。

Note

必须将 T2 实例启动到 VPC 的子网中。如果选择 `t2.micro` 实例，但是没有 VPC，系统会为您创建一个。此 VPC 在区域的每个可用区中均包含一个公有子网。

7. 在 Configure Details (配置实例详细信息) 页面中，执行以下操作：
 - a. 对于 Name，为启动配置键入一个名称。
 - b. (可选) 对于 Purchasing option (购买选项)，您可以请求 Spot 实例，并指定您愿意支付的最高实例每小时价格。有关更多信息，请参阅在 [Auto Scaling 组中启动 Spot 实例](#) (p. 59)。
 - c. (可选) 对于 IAM role，选择要与实例关联的角色。有关更多信息，请参阅在 [Amazon EC2 实例上运行的应用程序的 IAM 角色](#) (p. 147)。
 - d. (可选) 默认情况下，已为 Auto Scaling 实例启用基本监控。要为 Auto Scaling 实例启用详细监控，请选择 Enable CloudWatch detailed monitoring (启用 CloudWatch 详细监控)。
 - e. 对于 Advanced Details 和 IP Address Type，选择一个选项。要连接至 VPC 中的实例，您必须选择一个分配公有 IP 地址的选项。如果您要连接到实例，但是不确定您是否有默认 VPC，请选择 Assign a public IP address to every instance。
 - f. 选择 Skip to review。
8. 在 Review 页面上，选择 Edit security groups。按照说明操作来选择现有安全组，然后选择 Review。
9. 在 Review 页上选择 Create launch configuration。
10. 对于 Select an existing key pair or create a new key pair，选择所列的选项之一。选中确认复选框，然后选择 Create launch configuration。

Warning

如果您需要连接到实例，请不要选择在没有密钥对的情况下继续。

使用命令行创建启动配置

您可以使用以下任一命令：

- [create-launch-configuration](#) (AWS CLI)
- [New-ASLaunchConfiguration](#) (适用于 Windows PowerShell 的 AWS 工具)

使用 EC2 实例创建启动配置

Amazon EC2 Auto Scaling 提供了一个选项，可使用正在运行的 EC2 实例中的属性创建启动配置。

Tip

您可以[直接从 EC2 实例创建 Auto Scaling 组](#) (p. 46)。如果使用此功能，Amazon EC2 Auto Scaling 还自动为您创建启动配置。

如果指定的实例有启动配置当前不支持的属性，则由 Auto Scaling 组启动的实例可能不同于原始 EC2 实例。

从头创建启动配置和从现有 EC2 实例创建启动配置之间存在差异。如果从头创建启动配置，需要指定映像 ID、实例类型、可选资源（如存储设备）和可选设置（如监控）。从运行的实例中创建启动配置时，Amazon EC2 Auto Scaling 从指定的实例中派生启动配置的属性。属性也可能来自从中启动实例的 AMI 的块储存设备映射，从而忽略在启动后添加的任何其他块储存设备。

如果使用正在运行的实例创建启动配置，可以通过将以下属性指定为相同请求的一部分，来覆盖这些属性：AMI、块储存设备、密钥对、实例配置文件、实例类型、内核、监控、部署租用、虚拟磁盘、安全组、Spot 价格、用户数据、该实例是否有关联的公有 IP 地址，以及该实例是否经过了 EBS 优化。

以下示例介绍如何从 EC2 实例创建启动配置。

示例

- [使用 EC2 实例创建启动配置 \(p. 30\)](#)
- [从实例创建启动配置以及覆盖块储存设备 \(AWS CLI\) \(p. 31\)](#)
- [创建启动配置和覆盖实例类型 \(AWS CLI\) \(p. 32\)](#)

使用 EC2 实例创建启动配置

要使用现有 EC2 实例的属性创建启动配置，请指定该实例的 ID。

Important

用于启动指定实例的 AMI 必须仍然存在。

从 EC2 实例创建启动配置 (控制台)

您可以使用该控制台从运行中的 EC2 实例创建启动配置和 Auto Scaling 组，还可以将该实例添加到新的 Auto Scaling 组。有关更多信息，请参阅[将 EC2 实例附加到 Auto Scaling 组 \(p. 67\)](#)。

从 EC2 实例创建启动配置 (AWS CLI)

可以使用以下 `create-launch-configuration` 命令从一个实例中创建启动配置，它使用与该实例相同的属性。将忽略在启动后添加的任何块储存设备。

```
aws autoscaling create-launch-configuration --launch-configuration-name my-lc-from-instance
--instance-id i-a8e09d9c
```

您可以使用以下 `describe-launch-configurations` 命令描述启动配置，并确认其属性与实例的属性相匹配：

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-lc-from-
instance
```

以下为响应示例：

```
{
  "LaunchConfigurations": [
    {
      "UserData": null,
      "EbsOptimized": false,
      "LaunchConfigurationARN": "arn",
      "InstanceMonitoring": {
        "Enabled": false
      },
      "ImageId": "ami-05355a6c",
      "CreatedTime": "2014-12-29T16:14:50.382Z",
      "BlockDeviceMappings": [],
      "KeyName": "my-key-pair",
      "SecurityGroups": [
        "sg-8422d1eb"
      ],
      "LaunchConfigurationName": "my-lc-from-instance",
      "KernelId": "null",
      "RamdiskId": null,
      "InstanceType": "t1.micro",
      "AssociatePublicIpAddress": true
    }
  ]
}
```

```
}  
]  
}
```

从实例创建启动配置以及覆盖块储存设备 (AWS CLI)

默认情况下，Amazon EC2 Auto Scaling 使用您指定的 EC2 实例中的属性创建启动配置。不过，块储存设备来自用于启动实例的 AMI，而不是来自实例。要将块储存设备添加到启动配置，请覆盖该启动配置的块储存设备映射。

Important

用于启动指定实例的 AMI 必须仍然存在。

创建启动配置和覆盖块储存设备

可以使用以下 `create-launch-configuration` 命令创建一个启动配置，它使用 EC2 实例但具有自定义块储存设备映射：

```
aws autoscaling create-launch-configuration --launch-configuration-name my-lc-from-  
instance-bdm --instance-id i-a8e09d9c \  
--block-device-mappings "[{\"DeviceName\":\"/dev/sda1\",\"Ebs\":{\"SnapshotId\":  
\"snap-3decf207\"}}, {\"DeviceName\":\"/dev/sdf\",\"Ebs\":{\"SnapshotId\":\"snap-  
eed6ac86\"}}]"
```

可以使用以下 `describe-launch-configurations` 命令描述启动配置，并确认它使用自定义块储存设备映射：

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-lc-from-  
instance-bdm
```

下面的示例响应描述了该启动配置：

```
{  
  "LaunchConfigurations": [  
    {  
      "UserData": null,  
      "EbsOptimized": false,  
      "LaunchConfigurationARN": "arn",  
      "InstanceMonitoring": {  
        "Enabled": false  
      },  
      "ImageId": "ami-c49c0dac",  
      "CreatedTime": "2015-01-07T14:51:26.065Z",  
      "BlockDeviceMappings": [  
        {  
          "DeviceName": "/dev/sda1",  
          "Ebs": {  
            "SnapshotId": "snap-3decf207"  
          }  
        },  
        {  
          "DeviceName": "/dev/sdf",  
          "Ebs": {  
            "SnapshotId": "snap-eed6ac86"  
          }  
        }  
      ],  
      "KeyName": "my-key-pair",  
      "SecurityGroups": [  
        {  
          "GroupId": "sg-1a2b3c4d"  
        }  
      ]  
    }  
  ]  
}
```

```
        "sg-8637d3e3"
      ],
      "LaunchConfigurationName": "my-lc-from-instance-bdm",
      "KernelId": null,
      "RamdiskId": null,
      "InstanceType": "t1.micro",
      "AssociatePublicIpAddress": true
    }
  ]
}
```

创建启动配置和覆盖实例类型 (AWS CLI)

默认情况下，Amazon EC2 Auto Scaling 使用您指定的 EC2 实例中的属性创建启动配置。根据您的要求，您可能需要覆盖实例中的属性并使用所需的值。例如，您可以覆盖实例类型。

Important

用于启动指定实例的 AMI 必须仍然存在。

创建启动配置和覆盖实例类型

可以使用以下 [create-launch-configuration](#) 命令创建启动配置，它使用一个 EC2 实例但实例类型（例如 t2.medium）不同于该实例（例如 t2.micro）：

```
aws autoscaling create-launch-configuration --launch-configuration-name my-lc-from-instance-changetype \
--instance-id i-a8e09d9c --instance-type t2.medium
```

可以使用以下 [describe-launch-configurations](#) 命令描述启动配置，并确认覆盖了实例类型：

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-lc-from-instance-changetype
```

下面的示例响应描述了该启动配置：

```
{
  "LaunchConfigurations": [
    {
      "UserData": null,
      "EbsOptimized": false,
      "LaunchConfigurationARN": "arn",
      "InstanceMonitoring": {
        "Enabled": false
      },
      "ImageId": "ami-05355a6c",
      "CreatedTime": "2014-12-29T16:14:50.382Z",
      "BlockDeviceMappings": [],
      "KeyName": "my-key-pair",
      "SecurityGroups": [
        "sg-8422d1eb"
      ],
      "LaunchConfigurationName": "my-lc-from-instance-changetype",
      "KernelId": "null",
      "RamdiskId": null,
      "InstanceType": "t2.medium",
      "AssociatePublicIpAddress": true
    }
  ]
}
```

```
}
```

更改 Auto Scaling 组的启动配置

Auto Scaling 组一次与一个启动配置关联，并且启动配置在创建后不能修改。要更改 Auto Scaling 组的启动配置，请将现有的启动配置作为新启动配置的基础。然后，更新 Auto Scaling 组以使用新的启动配置。

更改 Auto Scaling 组的启动配置后，将使用新的配置选项启动所有新实例，但现有实例不受影响。

更改 Auto Scaling 组的启动配置 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Launch Configurations。
3. 选择启动配置，选择 Actions，然后单击 Copy launch configuration。这将设置与原启动配置选项相同的新启动配置，但在名称中会增加“Copy”文本。
4. 在 Copy Launch Configuration 页面上，根据需要编辑配置选项，然后选择 Create launch configuration。
5. 在确认页面上，选择 View your Auto Scaling groups (查看您的 Auto Scaling 组)。
6. 选择 Auto Scaling 组，然后依次选择 Details (详细信息) 和 Edit (编辑)。
7. 从 Launch Configuration 中选择新的启动配置，然后选择 Save。

更改 Auto Scaling 组的启动配置 (AWS CLI)

1. 使用 `describe-launch-configurations` 命令描述当前的启动配置。
2. 使用 `create-launch-configuration` 命令创建新的启动配置。
3. 使用 `update-auto-scaling-group` 命令和 `--launch-configuration-names` 参数更新 Auto Scaling 组的启动配置。

更改 Auto Scaling 组的启动配置 (Windows PowerShell 工具)

1. 使用 `Get-ASLaunchConfiguration` 命令描述当前的启动配置。
2. 使用 `New-ASLaunchConfiguration` 命令创建新的启动配置。
3. 使用 `Update-ASAutoScalingGroup` 命令和 `-LaunchConfigurationName` 参数更新 Auto Scaling 组的启动配置。

在 VPC 中启动 Auto Scaling 实例

通过使用 Amazon Virtual Private Cloud (Amazon VPC)，您可以在 AWS 云的私有隔离部分中定义虚拟联网环境。您的虚拟联网环境完全由您控制。有关更多信息，请参阅 [Amazon VPC 用户指南](#)。

在 Virtual Private Cloud (VPC) 内，您可启动 AWS 资源，如 Auto Scaling 组。Auto Scaling 组在 VPC 中的工作原理与它在 Amazon EC2 中的工作原理基本相同，支持同样的功能集。

Amazon VPC 中的子网是可用区的一个部分，由 VPC 的 IP 地址范围段定义。您可以根据您的安全和运行需求，使用子网对您的实例进行分组。子网完全位于其最初创建时所在的可用区内。在子网中启动 Auto Scaling 实例。

要支持 Internet 与您的子网中的实例之间的通信，您必须创建一个 Internet 网关并将其附加到您的 VPC。通过使用 Internet 网关，您的子网中的资源可以通过 Amazon EC2 网络边界连接到 Internet。如果子网的流量

路由到 Internet 网关，则子网称为公有子网。如果子网的流量不路由到 Internet 网关，则子网称为私有子网。请将公有子网用于必须连接到 Internet 的资源，并将私有子网用于不需要连接到 Internet 的资源。

先决条件

在 VPC 中启动 Auto Scaling 实例之前，您必须先创建 VPC 环境。创建了 VPC 和子网后，在子网内启动 Auto Scaling 实例。创建带单个公有子网的 VPC 的最简单方法就是使用 VPC 向导。有关更多信息，请参阅 [Amazon VPC 入门指南](#)。

目录

- [默认 VPC \(p. 34\)](#)
- [VPC 中的 IP 寻址 \(p. 34\)](#)
- [实例部署租期 \(p. 34\)](#)
- [将 EC2-Classic 实例链接到 VPC \(p. 35\)](#)

默认 VPC

如果您在 2013 年 12 月 4 日以后创建了 AWS 账户，或者您要在新区域中创建 Auto Scaling 组，则我们会为您创建默认 VPC。默认 VPC 随每个可用区中的默认子网提供。如果您有默认 VPC，则在默认情况下，将在默认 VPC 中创建 Auto Scaling 组。

有关默认 VPC 以及检查您的账户是否附带默认 VPC 的信息，请参阅 [Amazon VPC 开发人员指南](#) 中的 [您的默认 VPC 和子网](#)。

VPC 中的 IP 寻址

在 VPC 中启动 Auto Scaling 实例时，将自动为这些实例分配子网地址范围内的私有 IP 地址。这样，这些实例就能够与 VPC 中的其他实例通信。

您可以进行启动配置，为您的实例分配公有 IP 地址。通过将公有 IP 地址分配给实例，这些实例可以与 Internet 或 AWS 中的其他服务进行通信。

如果为实例启用公有 IP 地址，并将这些实例启动到配置为自动分配 IPv6 地址的子网，则这些实例接收 IPv4 和 IPv6 地址。否则，这些实例仅接收 IPv4 地址。有关更多信息，请参阅 [Amazon EC2 用户指南](#)（适用于 Linux 实例）中的 [IPv6 地址](#)。

实例部署租期

专用实例在主机硬件级别上与您的非专用实例和属于其他 AWS 账户的实例实现物理隔离。当您创建 VPC 时，默认情况下它的租期属性设置为 default。在这样一个 VPC 中，您可启动租期值为 dedicated 的实例，以便它们作为单一租期实例运行。否则，默认情况下，这些实例作为共享租期实例运行。如您将 VPC 的租期属性设置为 dedicated，则在 VPC 内启动的所有实例都作为单一租期实例运行。有关更多信息，请参阅 [Amazon VPC 用户指南](#) 中的 [专用实例](#)。有关定价信息，请参阅 [Amazon EC2 专用实例](#)。

在创建启动配置后，实例部署租期的默认值为 null，实例租期由 VPC 的租期属性控制。下表总结了在 VPC 中启动的 Auto Scaling 实例的实例部署租期。

启动配置租期	VPC 租期 = 默认	VPC 租期 = 专用
未指定	共享租期实例	专用实例
default	共享租期实例	专用实例
dedicated	专用实例	专用实例

您可以使用 `create-launch-configuration` 命令和 `--placement-tenancy` 选项，将启动配置的实例部署租期指定为 `default` 或 `dedicated`。例如，以下命令将启动配置租期设置为 `dedicated`：

```
aws autoscaling create-launch-configuration --launch-configuration-name my-launch-config --placement-tenancy dedicated --image-id ...
```

您可以使用以下 `describe-launch-configurations` 命令检查启动配置的实例部署租期：

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-launch-config
```

以下是创建专用实例的启动配置的示例输出。只有在明确设置实例部署租期时，`PlacementTenancy` 参数才是该命令输出的一部分。

```
{
  "LaunchConfigurations": [
    {
      "UserData": null,
      "EbsOptimized": false,
      "PlacementTenancy": "dedicated",
      "LaunchConfigurationARN": "arn",
      "InstanceMonitoring": {
        "Enabled": true
      },
      "ImageId": "ami-b5a7ea85",
      "CreatedTime": "2015-03-08T23:39:49.011Z",
      "BlockDeviceMappings": [],
      "KeyName": null,
      "SecurityGroups": [],
      "LaunchConfigurationName": "my-launch-config",
      "KernelId": null,
      "RamdiskId": null,
      "InstanceType": "m3.medium"
    }
  ]
}
```

将 EC2-Classic 实例链接到 VPC

如果您在 EC2-Classic 的 Auto Scaling 组中启动实例，您可使用 `ClassicLink` 将这些实例链接到 VPC。通过使用 `ClassicLink`，您可以将 VPC 的一个或多个安全组与 Auto Scaling 组中的 EC2-Classic 实例相关联。这样，这些链接的 EC2-Classic 实例就可以使用私有 IP 地址与 VPC 中的实例进行通信。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [ClassicLink](#)。

如果您的 Auto Scaling 组中存在运行的 EC2-Classic 实例，您可将它们链接到启用 `ClassicLink` 的 VPC。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [将实例链接到 VPC](#)。或者，您也可以更新 Auto Scaling 组以使用一个启动配置，该配置在启动时将 EC2-Classic 实例自动链接到一个 VPC。然后，终止运行的实例，并让 Auto Scaling 组启动链接到 VPC 的新实例。

链接到 VPC（控制台）

使用以下程序创建启动配置，通过该启动配置将 EC2-Classic 实例链接到指定的 VPC，并更新现有 Auto Scaling 组来使用启动配置。

将 Auto Scaling 组中的 EC2-Classic 实例链接到 VPC。

1. 确认 VPC 已启用 `ClassicLink`。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [查看启用了 ClassicLink 的 VPC](#)。

2. 为要将 EC2-Classical 实例链接到的 VPC 创建一个安全组。添加规则以控制链接的 EC2-Classical 实例和 VPC 中的实例之间的通信。
3. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
4. 在导航窗格上，选择 Launch Configurations。
5. 选择 Create launch configuration。
6. 在 Choose AMI 页面中，选择 AMI。
7. 在 Choose an Instance Type 页面上，选择实例类型，然后选择 Next: Configure details。
8. 在 Configure details 页面上，执行以下操作：
 - a. 为您的启动配置键入一个名称。
 - b. 展开 Advanced Details，选择您需要的 IP Address Type，然后选择 Link to VPC。
 - c. 对于 VPC，选择在步骤 1 中启用了 ClassicLink 的 VPC。
 - d. 对于 Security Groups，选择在步骤 2 中创建的安全组。
 - e. 选择 Skip to review。
9. 在 Review 页面上，进行所需的更改，然后选择 Create launch configuration。对于 Select an existing key pair or create a new key pair，选择一个选项，选中确认复选框（如果出现），然后选择 Create launch configuration。
10. 当出现提示时，请遵循说明创建使用新启动配置的 Auto Scaling 组。务必为 Network 选择 Launch into EC2-Classical。否则，请选择 Cancel (取消) 并将您的启动配置添加到一个现有 Auto Scaling 组，如下所示：
 - a. 在导航窗格上，选择 Auto Scaling Groups。
 - b. 选择您的 Auto Scaling 组，然后选择 Actions (操作) 和 Edit (编辑)。
 - c. 对于 Launch Configuration，选择新的启动配置，然后选择 Save。

链接到 VPC (AWS CLI)

使用以下程序创建启动配置，通过该启动配置将 EC2-Classical 实例链接到指定的 VPC，并更新现有 Auto Scaling 组来使用该启动配置。

将 Auto Scaling 组中的 EC2-Classical 实例链接到 VPC

1. 确认 VPC 已启用 ClassicLink。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[查看启用了 ClassicLink 的 VPC](#)。
2. 为要将 EC2-Classical 实例链接到的 VPC 创建一个安全组。添加规则以控制链接的 EC2-Classical 实例和 VPC 中的实例之间的通信。
3. 使用 `create-launch-configuration` 命令创建一个启动配置，如下所示。为 `vpc_id` 指定一个值以作为步骤 1 中启用了 ClassicLink 的 VPC 的 ID，并为 `group_id` 指定一个值以作为步骤 2 中的安全组：

```
aws autoscaling create-launch-configuration --launch-configuration-name classiclink-config \
  --image-id ami_id --instance-type instance_type \
  --classic-link-vpc-id vpc_id --classic-link-vpc-security-groups group_id
```

4. 使用在上一步中创建的启动配置来更新您的现有 Auto Scaling 组，例如 `my-asg`。这一 Auto Scaling 组中启动的任何新 EC2-Classical 实例都是已链接的 EC2-Classical 实例。按以下方式使用 `update-auto-scaling-group` 命令：

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \
  --launch-configuration-name classiclink-config
```

或者，您也可以将该启动配置与使用 `create-auto-scaling-group` 创建的新 Auto Scaling 组一起使用。

Auto Scaling 组

Auto Scaling 组包含 Amazon EC2 实例集合，这些集合被视为逻辑组以用于自动扩展和管理用途。另外，Auto Scaling 组让您能够使用 Amazon EC2 Auto Scaling 功能，如运行状况检查替换和扩展策略。保持 Auto Scaling 组中的实例数量和自动扩展都是 Amazon EC2 Auto Scaling 服务的核心功能。

Auto Scaling 组的大小取决于您设置为所需容量的实例数量。您可以通过手动方式或使用自动扩展调整其大小以满足需求。

Auto Scaling 组会首先启动足够实例以达到需要的容量。它通过对组中实例定期执行运行状况检查来保持实例数量。即使某个实例运行状况不佳，Auto Scaling 组也会继续保持固定数量的实例。如果某个实例运行状况不佳，则该组终止运行状况不佳的实例，并启动其他实例来替换它。有关更多信息，请参阅[Auto Scaling 实例的运行状况检查 \(p. 115\)](#)。

您可以使用扩展策略动态增加或减少组中的实例数量，以符合不断变化的条件。扩展策略生效时，Auto Scaling 组会在您指定的最小和最大容量值之间调整组的所需容量，并根据需要启动或终止实例。您还可以按计划进行扩展。有关更多信息，请参阅[扩展 Auto Scaling 组的大小 \(p. 64\)](#)。

Auto Scaling 组可以启动按需实例和/或 Spot 实例。仅在您将 Auto Scaling 组配置为使用启动模板时，才可以为该组指定多个购买选项。（我们建议您使用启动模板而不是启动配置，以确保您可以使用 Amazon EC2 的最新功能。）

Spot 实例可为您提供与按需价格相关的极低折扣，获得未使用的 Amazon EC2 容量。有关更多信息，请参阅[Amazon EC2 Spot 实例](#)。Spot 实例和按需实例的主要差别在于，Spot 实例价格会根据需求变动，并且 Amazon EC2 可根据 Spot 实例可用情况或价格的变化终止单个 Spot 实例。Spot 实例终止后，Auto Scaling 组会尝试启动替代实例来保持该组的所需容量。

实例启动时，如果您指定多个可用区，会为这些可用区分配所需容量。如果执行扩展操作，Amazon EC2 Auto Scaling 会自动保持您指定的所有可用区的平衡。

本节提供创建 Auto Scaling 组的详细步骤。如果您刚开始使用 Auto Scaling 组，从[Amazon EC2 Auto Scaling 入门 \(p. 11\)](#)开始以了解 Amazon EC2 Auto Scaling 中使用的各个构建块。

目录

- [Auto Scaling 组具有多个实例类型和购买选项 \(p. 37\)](#)
- [使用启动模板创建 Auto Scaling 组 \(p. 44\)](#)
- [使用启动配置创建 Auto Scaling 组 \(p. 45\)](#)
- [使用 EC2 实例创建 Auto Scaling 组 \(p. 46\)](#)
- [使用 Amazon EC2 启动向导创建 Auto Scaling 组 \(p. 48\)](#)
- [标记 Auto Scaling 组和实例 \(p. 49\)](#)
- [将负载均衡器与 Auto Scaling 组结合使用 \(p. 53\)](#)
- [在 Auto Scaling 组中启动 Spot 实例 \(p. 59\)](#)
- [将 Auto Scaling 组合并为单个多区组 \(p. 59\)](#)
- [删除 Auto Scaling 基础设施 \(p. 61\)](#)

Auto Scaling 组具有多个实例类型和购买选项

您可以启动并自动扩展单个 Auto Scaling 组中的一组按需实例和 Spot 实例。除接收折扣以使用 Spot 实例外，如果您指定存在与其匹配的预留实例的实例类型，则常规按需实例定价的折扣费率同样适用。按需实

例与预留实例之间唯一的差异是您必须提前购买预留实例。以上所有因素的综合作用是帮助您进一步节约 Amazon EC2 实例成本，同时确保为您的应用程序获得所需的规模和性能。

首先在一个启动模板中指定常用配置参数，并在创建 Auto Scaling 组时选择该模板。在配置 Auto Scaling 组时，您可以指定要启动多少按需容量和 Spot 容量，选择最适合您的应用程序的实例类型（覆盖启动模板指定的实例类型），并在每个购买选项中定义 Amazon EC2 Auto Scaling 应如何跨实例类型分配您的容量。

通过跨多个可用区中运行的多个实例类型部署应用程序以提高可用性。您必须指定至少两个实例类型，但最佳实践是选择几个实例类型以避免从容量不足的实例池中启动实例。如果一个 Spot 实例池无法满足 Auto Scaling 组的 Spot 实例请求，它将继续尝试其他 Spot 实例池而不是启动按需实例，使您可以利用 Spot 实例节约成本。

分配策略

以下分配策略决定 Auto Scaling 组如何根据可用的实例类型满足按需实例和 Spot 容量。

在每种情况下，Amazon EC2 Auto Scaling 首先尝试确保实例在您指定的可用区之间均匀平衡。然后，它根据指定的分配策略启动实例类型。

按需实例

按需实例的分配策略是 prioritized。Amazon EC2 Auto Scaling 使用启动模板覆盖列表中的实例类型顺序，确定在满足按需容量时应首先使用哪一实例类型。

例如，您按以下顺序指定三个启动模板覆盖：c5.large、c4.large 和 c3.large。在启动您的按需实例时，Auto Scaling 组满足按需容量的顺序是从 c5.large 开始，然后是 c4.large，再然后是 c3.large。如果存在未使用的 c4.large 预留实例，您可以设置实例类型的优先级，通过指定 c4.large 作为最高优先级的实例类型来使您的预留实例具有最高优先级。当 c4.large 实例启动时，您可享受预留实例定价。

Spot 实例

Amazon EC2 Auto Scaling 提供了两种类型的分配策略可用于 Spot 实例：

capacity-optimized

Amazon EC2 Auto Scaling 可以从 Spot 实例池中分配实例，并为启动的实例数量提供最佳容量。通过这种方式部署可帮助您最高效地使用剩余 EC2 容量。

使用 Spot 实例，定价会根据长期供需趋势缓慢发生变化，但容量会实时波动。该 capacity-optimized 策略通过查看实时容量数据并预测可用性最高的池，自动在可用性最高的池中启动 Spot 实例。这适用于与中断相关的重启工作和检查点成本较高的工作负载，例如大数据和分析、图像和媒体渲染、机器学习以及高性能计算。通过实现更低的中断可能性，capacity-optimized 策略可以降低您工作负载的整体成本。

lowest-price

Amazon EC2 Auto Scaling 从您指定数量 (N) 的 Spot 实例池以及在满足需求时单价最低的池中分配您的实例。

例如，如果您指定 4 个实例类型和 4 个可用区，您的 Auto Scaling 组有可能利用多达 16 个不同的 Spot 实例池。如果您为分配策略指定 2 个 Spot 池 (N = 2)，您的 Auto Scaling 组最少可以使用 8 个不同池中单价最低的 Spot 池来满足 Spot 容量。

首先，我们建议您选择 capacity-optimized 分配策略并指定适合您的应用程序的若干种实例类型。此外，您可以为 Amazon EC2 Auto Scaling 定义一系列可用区，以便在启动实例时从中选择。

（可选）您可以指定您的 Spot 实例的最高价。如果未指定最高价，则默认最高价为按需价格，不过您仍可以获得 Spot 实例提供的大幅折扣。这些折扣得以实现是得益于使用新的 [Spot 定价模型](#) 带来了稳定的 Spot 定价。

控制按需实例的比例

您可以完全控制 Auto Scaling 组中作为按需实例启动的实例所占比例。为确保您始终拥有实例容量，您可以指定组中作为按需实例启动的百分比，也可以选择指定初始的按需实例基数。如果您选择指定按需实例容量基本容量，当 Auto Scaling 扩展时，会确保首先启动按需实例的该基本容量。超出基本容量的任何内容都使用按需百分比来确定要启动多少按需实例和 Spot 实例。您可以指定 0 到 100 之间的任意数字来作为按需百分比。

在 Auto Scaling 组大小增加时，该组的行为如下所示：

示例：扩展行为

实例分配	各购买选项的运行实例总数			
	10	20	30	40
示例 1				
按需基准：10	10	10	10	10
超过基准的按需百分比：50%	0	5	10	15
Spot 百分比：50%	0	5	10	15
示例 2				
按需基准：0	0	0	0	0
超过基准的按需百分比：0%	0	0	0	0
Spot 百分比：100%	10	20	30	40
示例 3				
按需基准：0	0	0	0	0
超过基准的按需百分比：60%	6	12	18	24
Spot 百分比：40%	4	8	12	16
示例 4				
按需基准：0	0	0	0	0
超过基准的按需百分比：100%	10	20	30	40
Spot 百分比：0%	0	0	0	0
示例 5				
按需基准：12	10	12	12	12
超过基准的按需百分比：0%	0	0	0	0

实例分配	各购买选项的运行实例总数			
Spot 百分比： 100%	0	8	18	28

Spot 实例的最佳实践

创建 Auto Scaling 组以请求 Spot 实例之前，请查看 [Spot 最佳实践](#)。使用这些最佳实践规划您的请求，以便以可能的最低价格预置需要的实例类型。还建议执行以下操作：

- 使用默认最高价，这是按需价格。您只需为您启动的 Spot 实例支付 Spot 市场价格。如果 Spot 市场价格在最高价范围内，则是否满足您的请求取决于可用性。有关更多信息，请参阅适用于 Linux 实例的 Amazon EC2 用户指南中的 [定价和节约成本](#)。
- 创建包括多种实例类型的 Auto Scaling 组。最少包括两种。由于可用区内每个实例类型的价格分开波动，因此，如果您灵活使用实例类型，则通常可以在价格不变的情况下获得更多计算容量。
- 同样，不要局限为仅选择最受欢迎的实例类型。由于价格是根据长期需求进行调整的，常见实例类型（如最近启动的实例系列）的价格调整频率往往更高。选择不太热门的较旧实例类型往往能降低成本和减少中断。
- 如果您选择 lowest-price 分配策略并运行 Web 服务，则需要指定较大的 Spot 池数量（例如 N=10），以减少其中一个可用区中的池暂时不可用时产生的 Spot 实例中断影响。如果您运行批处理或其他非关键任务应用程序，可以指定较小 Spot 池数量，例如，N = 2。这有助于确保您通过每个可用区中价格最低的 Spot 池提供 Spot 实例。
- 使用 Spot 实例中断通知监控您的 Spot 实例的状态。例如，您可以在 Amazon CloudWatch Events 中设置一个规则，自动将 EC2 Spot 两分钟警告发送到 Amazon SNS 主题、AWS Lambda 函数或其他目标。有关更多信息，请参阅适用于 Linux 实例的 Amazon EC2 用户指南和 [Amazon CloudWatch Events 用户指南](#) 中的 [Spot 实例中断通知](#)。

先决条件

您的启动模板配置为使用 Auto Scaling 组。有关信息，请参阅 [为 Auto Scaling 组创建启动模板 \(p. 22\)](#)。

IAM 用户仅在具有调用 `ec2:RunInstances` 操作的权限时，才能使用启动模板创建 Auto Scaling 组。有关配置用户权限的信息，请参阅 [控制对您 Amazon EC2 Auto Scaling 资源的访问权 \(p. 134\)](#)。

使用多个购买选项创建 Auto Scaling 组

请按照以下步骤创建您可以扩展的按需实例和 Spot 实例队组。您可以使用 Instances Distribution (实例分布) 中的默认设置快速创建队组。否则，您可以修改任意默认设置。您对组扩展拥有完全控制权。您可以将所需容量设置为特定的实例数量，或者您可以利用自动扩展以根据实际需求调整容量。

使用多个购买选项创建 Auto Scaling 组（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏中，选择在创建启动模板时使用的同一 AWS 区域。
3. 在导航窗格中，选择 Auto Scaling Groups。
4. 选择 Create Auto Scaling group (创建 Auto Scaling 组)。
5. 选择启动模板，选择您的启动模板，然后选择下一步。
6. 在配置 Auto Scaling 组详细信息页上，为组名输入 Auto Scaling 组的名称。
7. 对于 Launch template version (启动模板版本)，选择 Auto Scaling 组在扩展时使用启动模板的默认版本、最新版本还是特定版本。

8. 为 Fleet Composition (队组组合) 选择 Combine purchase options and instances (组合购买选项和实例), 以使用按需和 Spot 购买选项跨多个实例类型启动实例。
9. 为实例类型选择可以启动的最佳实例类型 (如 c5.large)。您添加实例类型时的顺序设定了它们对于按需实例的优先级。当 Auto Scaling 组启动按需容量时, 位于列表顶部的实例类型优先级最高。您必须指定至少两个实例类型。
10. 对于 Instances Distribution (实例分布), 选择是要保留还是编辑默认设置。清除 Use the default settings to get started quickly (使用默认设置以快速入门), 即可编辑默认设置。
11. 如果您选择编辑默认设置, 请提供以下信息。
 - 对于最高 Spot 价格, 选择使用默认设置以将最高 Spot 价格限定为按需价格。或者选择 Set your maximum price (设置最高价) 并输入值, 以指定您愿意为 Spot 实例支付的、按小时计算的实例价格。
 - 对于 Spot Allocation Strategy (Spot 分配策略), 默认值为 Launch Spot Instances optimally based on the available Spot capacity per Availability Zone (根据各可用区的可用 Spot 容量启动最优 Spot 实例)。要改为使用最低价格策略, 请选择 Diversify Spot Instances across your N lowest priced instance types per Availability Zone (使各可用区的 N 个最低价格实例类型的 Spot 实例实现多样化), 然后输入要使用的 Spot 池数量, 或者接受默认值 (2)。
 - 对于 Optional On-Demand Base (可选的按需基准), 指定必须使用按需实例满足的、Auto Scaling 组的初始容量的最小实例数。
 - 对于 On-Demand Percentage Above Base (超过基准的按需百分比), 为超出可选按需基准数量的额外容量指定按需实例和 Spot 实例的百分比。
12. 对于组大小, 输入 Auto Scaling 组的初始实例数。
13. 对于 Network (网络), 为您的 Auto Scaling 组选择 VPC。

Note

EC2-Classic 中不支持使用实例类型和购买选项的组合来启动实例。

14. 对于 Subnet (子网), 选择指定 VPC 中的一个或多个子网。可以在多个可用区中使用子网以提高高可用性。有关 Amazon EC2 Auto Scaling 高可用性的更多信息, 请参阅[在可用区之间分配实例 \(p. 5\)](#)。
15. (可选) 要在负载均衡器中注册您的 Amazon EC2 实例, 依次选择高级详细信息和从一个或多个负载均衡器接收流量, 然后选择一个或多个 Classic Load Balancer 或目标组。
16. 选择 Next: Configure scaling policies。
17. 在 Configure scaling policies (配置扩展策略) 页面上, 选择下列选项之一, 然后选择 Next: Configure Notifications (下一步: 配置通知):
 - 要根据需要手动调整 Auto Scaling 组的大小, 请选择 Keep this group at its initial size (保留该组的初始大小)。有关更多信息, 请参阅[Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)。
 - 要根据您指定的条件自动调整 Auto Scaling 组的大小, 选择 Use scaling policies to adjust the capacity of this group (使用扩展处理调整该组的容量), 然后按照指示操作。有关详细信息, 请参阅[配置扩展策略 \(p. 79\)](#)。
18. (可选) 要接收通知, 请选择 Add notification, 配置通知, 然后选择 Next: Configure Tags。
19. (可选) 要添加标签, 请选择 Edit tags, 为每个标签提供标签键和值, 然后选择 Review。

或者, 您可以稍后添加标签。有关更多信息, 请参阅[标记 Auto Scaling 组和实例 \(p. 49\)](#)。

20. 在 Review 页面上, 选择 Create Auto Scaling group。
21. 在 Auto Scaling group creation status 页面上, 选择 Close。

使用多个购买选项创建 Auto Scaling 组 (AWS CLI)

以下 `create-auto-scaling-group` 命令会创建一个 Auto Scaling 组, 该组指定以下内容:

- 实例类型按优先级顺序启动 (c5.large、c4.large、c3.large)
- 要为其分配 Spot 实例的 Spot 实例池 (2)
- 要在其中启动实例的子网 (subnet-5ea0c127、subnet-6194ea3b、subnet-c934b782)


```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

下面是一个 config.json 示例文件。

```
{
  "AutoScalingGroupName": "my-asg",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {
      "LaunchTemplateSpecification": {
        "LaunchTemplateName": "my-template-for-auto-scaling",
        "Version": "1"
      },
      "Overrides": [
        {
          "InstanceType": "c5.large"
        },
        {
          "InstanceType": "c4.large"
        },
        {
          "InstanceType": "c3.large"
        }
      ]
    },
    "InstancesDistribution": {
      "OnDemandBaseCapacity": 1,
      "OnDemandPercentageAboveBaseCapacity": 50,
      "SpotInstancePools": 2
    }
  },
  "MinSize": 1,
  "MaxSize": 5,
  "DesiredCapacity": 3,
  "VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",
  "Tags": []
}
```

要验证该组是否已启动实例

使用以下 `describe-auto-scaling-groups` 命令：

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg
```

以下示例响应显示所需的容量为 3 并且组具有三个运行的实例。

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupARN": "arn",
      "ServiceLinkedRoleARN": "arn",
      "TargetGroupARNs": [],
      "SuspendedProcesses": [],
      "DesiredCapacity": 3,
      "MixedInstancesPolicy": {
        "InstancesDistribution": {
          "SpotAllocationStrategy": "lowest-price",
          "OnDemandPercentageAboveBaseCapacity": 50,
          "OnDemandAllocationStrategy": "prioritized",
          "SpotInstancePools": 2,
          "OnDemandBaseCapacity": 1
        }
      }
    }
  ]
}
```

```
"LaunchTemplate": {
  "LaunchTemplateSpecification": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "1",
    "LaunchTemplateId": "lt-050555ad16a3f9c7f"
  },
  "Overrides": [
    {
      "InstanceType": "c5.large"
    },
    {
      "InstanceType": "c4.large"
    },
    {
      "InstanceType": "c3.large"
    }
  ]
},
"EnabledMetrics": [],
"Tags": [],
"AutoScalingGroupName": "my-asg",
"DefaultCooldown": 300,
"MinSize": 1,
"Instances": [
  {
    "ProtectedFromScaleIn": false,
    "AvailabilityZone": "us-west-2a",
    "LaunchTemplate": {
      "LaunchTemplateName": "my-template-for-auto-scaling",
      "Version": "1",
      "LaunchTemplateId": "lt-050555ad16a3f9c7f"
    },
    "InstanceId": "i-0aae8709d49eeba4f",
    "HealthStatus": "Healthy",
    "LifecycleState": "InService"
  },
  {
    "ProtectedFromScaleIn": false,
    "AvailabilityZone": "us-west-2b",
    "LaunchTemplate": {
      "LaunchTemplateName": "my-template-for-auto-scaling",
      "Version": "1",
      "LaunchTemplateId": "lt-050555ad16a3f9c7f"
    },
    "InstanceId": "i-0c43f6003841d2d2b",
    "HealthStatus": "Healthy",
    "LifecycleState": "InService"
  },
  {
    "ProtectedFromScaleIn": false,
    "AvailabilityZone": "us-west-2c",
    "LaunchTemplate": {
      "LaunchTemplateName": "my-template-for-auto-scaling",
      "Version": "1",
      "LaunchTemplateId": "lt-050555ad16a3f9c7f"
    },
    "InstanceId": "i-0feb4cd6677d39903",
    "HealthStatus": "Healthy",
    "LifecycleState": "InService"
  }
],
"MaxSize": 5,
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",
"HealthCheckGracePeriod": 0,
"TerminationPolicies": [
```

```
        "Default"
      ],
      "LoadBalancerNames": [],
      "CreateTime": "2019-02-17T02:29:12.853Z",
      "AvailabilityZones": [
        "us-west-2a",
        "us-west-2b",
        "us-west-2c"
      ],
      "HealthCheckType": "EC2",
      "NewInstancesProtectedFromScaleIn": false
    }
  ]
}
```

使用启动模板创建 Auto Scaling 组

在创建 Auto Scaling 组时，您必须指定配置 Amazon EC2 实例所需的信息、实例的子网以及初始实例数。

要配置 Amazon EC2 实例，您可以指定启动配置、启动模板或 EC2 实例。以下过程演示如何使用启动模板创建 Auto Scaling 组。

使用启动模板时，您需要选择启动模板以及组启动 EC2 实例时所使用的启动模板的特定版本。您可以随时更新组来更改这些选择。或者，您可以将 Auto Scaling 组配置为在扩展事件发生时动态选择启动模板的默认版本或最新版本。例如，您可以配置 Auto Scaling 组以动态选择启动模板的默认版本。要更改通过组启动的 EC2 实例的配置，请创建或指定启动模板的新默认版本。

先决条件

- 您必须已创建启动模板，其中包含启动 EC2 实例所需的参数。有关这些参数以及在创建用于 Auto Scaling 组的启动模板时适用的限制的信息，请参阅[Auto Scaling 组创建启动模板 \(p. 22\)](#)。
- 您必须拥有 IAM 权限才能使用启动模板创建 Auto Scaling 组并为这些实例创建 EC2 资源。有关更多信息，请参阅[控制对您 Amazon EC2 Auto Scaling 资源的访问权 \(p. 134\)](#)。

使用启动模板创建 Auto Scaling 组 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏中，选择在创建启动模板时使用的同一 AWS 区域。
3. 在导航窗格中，选择 Auto Scaling Groups。
4. 选择 Create Auto Scaling group (创建 Auto Scaling 组)。
5. 选择启动模板，选择您的启动模板，然后选择下一步。
6. 在配置 Auto Scaling 组详细信息页上，为组名输入 Auto Scaling 组的名称。
7. 对于 Launch template version (启动模板版本)，选择 Auto Scaling 组在扩展时使用启动模板的默认版本、最新版本还是特定版本。
8. 对于队列构成，请选择遵循启动模板以使用在启动模板中指定的 EC2 实例类型和购买选项。

Note

或者，选择 Combine purchase options and instances (组合购买选项和实例)，以使用按需和 Spot 购买选项跨多个实例类型启动实例。有关更多信息，请参阅[Auto Scaling 组具有多个实例类型和购买选项 \(p. 37\)](#)。

9. 对于组大小，输入 Auto Scaling 组的初始实例数。
10. 对于 Network (网络)，为您的 Auto Scaling 组选择 VPC。

11. 对于 Subnet (子网)，选择指定 VPC 中的一个或多个子网。可以在多个可用区中使用子网以提供高可用性。有关 Amazon EC2 Auto Scaling 高可用性的更多信息，请参阅[在可用区之间分配实例 \(p. 5\)](#)。
12. (可选) 要在负载均衡器中注册您的 Amazon EC2 实例，依次选择高级详细信息和从一个或多个负载均衡器接收流量，然后选择一个或多个 Classic Load Balancer 或目标组。
13. 选择 Next: Configure scaling policies。
14. 在 Configure scaling policies (配置扩展策略) 页面上，选择下列选项之一，然后选择 Next: Configure Notifications (下一步：配置通知)：
 - 要根据需要手动调整 Auto Scaling 组的大小，请选择 Keep this group at its initial size (保留该组的初始大小)。有关更多信息，请参阅[Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)。
 - 要根据您指定的条件自动调整 Auto Scaling 组的大小，选择 Use scaling policies to adjust the capacity of this group (使用扩展处理调整该组的容量)，然后按照指示操作。有关详细信息，请参阅[配置扩展策略 \(p. 79\)](#)。
15. (可选) 要接收通知，请选择 Add notification，配置通知，然后选择 Next: Configure Tags。
16. (可选) 要添加标签，请选择 Edit tags，为每个标签提供标签键和值，然后选择 Review。

或者，您可以稍后添加标签。有关更多信息，请参阅[标记 Auto Scaling 组和实例 \(p. 49\)](#)。
17. 在 Review 页面上，选择 Create Auto Scaling group。
18. 在 Auto Scaling group creation status 页面上，选择 Close。

使用命令行创建 Auto Scaling 组

您可以使用以下任一命令：

- [create-auto-scaling-group](#) (AWS CLI)
- [New-ASAutoScalingGroup](#) (适用于 Windows PowerShell 的 AWS 工具)

使用启动配置创建 Auto Scaling 组

在创建 Auto Scaling 组时，您必须指定配置 Amazon EC2 实例所需的信息、实例的子网以及初始实例数。

Important

要配置 Amazon EC2 实例，您可以指定启动模板、启动配置或 EC2 实例。建议您使用启动模板以确保您可以使用 Amazon EC2 的最新功能。有关更多信息，请参阅[启动模板 \(p. 22\)](#)。

以下过程演示如何使用启动配置创建 Auto Scaling 组。您无法在创建启动配置后进行修改，但可以替换 Auto Scaling 组的启动配置。有关更多信息，请参阅[更改 Auto Scaling 组的启动配置 \(p. 33\)](#)。

先决条件

创建启动配置。有关更多信息，请参阅[创建启动配置 \(p. 28\)](#)。

使用启动配置创建 Auto Scaling 组 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在屏幕顶部的导航栏中，选择在创建启动配置时使用的同一 AWS 区域。
3. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
4. 选择 Create Auto Scaling group (创建 Auto Scaling 组)。
5. 在 Create Auto Scaling Group (创建 Auto Scaling 组) 页面上，选择 Launch Configuration (启动配置)，选择现有启动配置，然后选择下一步。

Note

如果您没有任何启动配置，系统首先会提示您创建一个启动配置，然后才能继续创建 Auto Scaling 组的步骤。

6. 在配置 Auto Scaling 组详细信息页面上，执行以下操作：
 - a. 对于组名，输入 Auto Scaling 组的名称。
 - b. 对于组大小，输入 Auto Scaling 组的初始实例数。
 - c. 对于 Network (网络)，为您的 Auto Scaling 组选择 VPC。
 - d. 对于子网，选择 VPC 中的一个或多个子网。可以在多个可用区中使用子网以提高可用性。有关 Amazon EC2 Auto Scaling 高可用性的更多信息，请参阅[在可用区之间分配实例 \(p. 5\)](#)。
 - e. (可选) 要在负载均衡器中注册您的 Amazon EC2 实例，依次选择高级详细信息和从一个或多个负载均衡器接收流量，然后选择一个或多个 Classic Load Balancer 或目标组。
 - f. 选择 Next: Configure scaling policies。
7. 在 Configure scaling policies (配置扩展策略) 页面上，选择下列选项之一，然后选择 Next: Configure Notifications (下一步：配置通知)：
 - 要根据需要手动调整 Auto Scaling 组的大小，请选择 Keep this group at its initial size (保留该组的初始大小)。有关更多信息，请参阅[Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)。
 - 要根据您指定的条件自动调整 Auto Scaling 组的大小，选择 Use scaling policies to adjust the capacity of this group (使用扩展处理调整该组的容量)，然后按照指示操作。有关详细信息，请参阅[配置扩展策略 \(p. 79\)](#)。
8. (可选) 要接收通知，请选择 Add notification，配置通知，然后选择 Next: Configure Tags。
9. (可选) 要添加标签，请选择 Edit tags，为每个标签提供标签键和值，然后选择 Review。

或者，您可以稍后添加标签。有关更多信息，请参阅[标记 Auto Scaling 组和实例 \(p. 49\)](#)。
10. 在 Review 页面上，选择 Create Auto Scaling group。
11. 在 Auto Scaling group creation status 页面上，选择 Close。

使用命令行创建 Auto Scaling 组

您可以使用以下任一命令：

- [create-auto-scaling-group](#) (AWS CLI)
- [New-ASAutoScalingGroup](#) (适用于 Windows PowerShell 的 AWS 工具)

使用 EC2 实例创建 Auto Scaling 组

在创建 Auto Scaling 组时，您必须指定配置 Amazon EC2 实例所需的信息、实例的子网以及初始实例数。

要配置 Amazon EC2 实例，您可以指定启动配置、启动模板或 EC2 实例。以下过程演示如何使用 EC2 实例创建 Auto Scaling 组。要使用启动配置或启动模板，请参阅[使用启动配置创建 Auto Scaling 组 \(p. 45\)](#)或[使用启动模板创建 Auto Scaling 组 \(p. 44\)](#)。

当您使用 EC2 实例创建 Auto Scaling 组时，Amazon EC2 Auto Scaling 会创建启动配置并将其与 Auto Scaling 组关联。该启动配置的名称与 Auto Scaling 组相同，同时它从指定的实例派生其属性，如 AMI ID、实例类型和可用区。

限制

以下是使用 EC2 实例创建 Auto Scaling 组的限制条件：

- 如果识别的实例有标签，标签不会复制到新 Auto Scaling 组的 Tags 属性。

- Auto Scaling 组包括 AMI 的块储存设备映射，用来启动实例；不包括实例启动后挂载的任何块储存设备。
- 如果识别的实例在多个负载均衡器中注册，则有关负载均衡器的信息不会复制到负载均衡器或新 Auto Scaling 组的目标组属性。

先决条件

在开始之前，请使用 Amazon EC2 控制台或 [describe-instances](#) 命令 (AWS CLI) 查找 EC2 实例的 ID。EC2 实例必须符合以下标准：

- 实例位于要创建 Auto Scaling 组的可用区中。
- 实例不是其他 Auto Scaling 组的成员。
- 实例处于 `running` 状态。
- 用于启动实例的 AMI 必须仍然存在。

目录

- [使用 EC2 实例创建 Auto Scaling 组 \(控制台 \)](#) (p. 47)
- [使用 EC2 实例创建 Auto Scaling 组 \(AWS CLI\)](#) (p. 47)

使用 EC2 实例创建 Auto Scaling 组 (控制台)

您可以使用该控制台从运行中的 EC2 实例创建一个 Auto Scaling 组并将该实例添加到新的 Auto Scaling 组。有关更多信息，请参阅[将 EC2 实例附加到 Auto Scaling 组](#) (p. 67)。

使用 EC2 实例创建 Auto Scaling 组 (AWS CLI)

可以使用以下 `create-auto-scaling-group` 命令从 EC2 实例 `i-7f12e649` 中创建 Auto Scaling 组 `my-asg-from-instance`。

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg-from-instance \
  --instance-id i-7f12e649 --min-size 1 --max-size 2 --desired-capacity 2
```

使用以下 `describe-auto-scaling-groups` 命令描述 Auto Scaling 组。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg-from-instance
```

如以下示例响应所示，该组的所需容量为 2，有 2 个运行实例，启动配置同样命名为 `my-asg-from-instance`。

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupARN": "arn",
      "HealthCheckGracePeriod": 0,
      "SuspendedProcesses": [],
      "DesiredCapacity": 2,
      "Tags": [],
      "EnabledMetrics": [],
      "LoadBalancerNames": [],
      "AutoScalingGroupName": "my-asg-from-instance",
      "DefaultCooldown": 300,
      "MinSize": 1,
      "Instances": [
        {
          "InstanceId": "i-6bd79d87",
```

```
        "AvailabilityZone": "us-west-2a",
        "HealthStatus": "Healthy",
        "LifecycleState": "InService",
        "LaunchConfigurationName": "my-asg-from-instance"
    },
    {
        "InstanceId": "i-6cd79d80",
        "AvailabilityZone": "us-west-2a",
        "HealthStatus": "Healthy",
        "LifecycleState": "InService",
        "LaunchConfigurationName": "my-asg-from-instance"
    }
],
"MaxSize": 2,
"VPCZoneIdentifier": "subnet-6bea5f06",
"TerminationPolicies": [
    "Default"
],
"LaunchConfigurationName": "my-asg-from-instance",
"CreatedTime": "2014-12-29T16:14:50.397Z",
"AvailabilityZones": [
    "us-west-2a"
],
"HealthCheckType": "EC2"
}
]
```

使用以下 `describe-launch-configurations` 命令描述启动配置 `my-asg-from-instance`。

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-asg-from-
instance
```

使用 Amazon EC2 启动向导创建 Auto Scaling 组

使用 Amazon EC2 启动向导，您可以在一个过程中创建启动配置和 Auto Scaling 组。如果要启动多个实例，并且希望从已在 Amazon EC2 启动向导中选择的设置创建新的启动配置和 Auto Scaling 组，这非常有用。您不能使用此选项使用现有启动配置创建 Auto Scaling 组。

使用启动向导创建启动配置和 Auto Scaling 组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在控制面板中，选择 Launch Instance。
3. 选择一个 AMI，然后在下一页面上选择一个实例类型，然后选择 Next: Configure Instance Details。
4. 在 Number of instances 中输入要启动的实例数，然后选择 Launch into Auto Scaling Group。无需在此页面上添加任何其他配置详细信息。
5. 在确认页面上，选择 Create Launch Configuration。
6. 您会切换到启动配置向导的步骤 3。已根据您在 Amazon EC2 启动向导中的选择选择了 AMI 和实例类型。为启动配置输入名称，根据需要配置所有其他设置，然后选择 Next: Add Storage。
7. 配置任何其他卷，然后选择 Next: Configure Security Group。
8. 创建一个新安全组或选择一个现有组，然后选择 Review。
9. 查看启动配置的详细信息，然后选择 Create launch configuration 来选择密钥对并创建启动配置。
10. 在 Configure Auto Scaling group details (配置 Auto Scaling 组详细信息) 页面中，已为您选中了您创建的启动配置，并用您在 Amazon EC2 启动向导中指定的实例数填写了 Group size (组大小)。为组输入名称，指定 VPC 和子网（如果需要），然后选择 Next: Configure scaling policies。
11. 在 Configure scaling policies 页面上，选择下列选项之一，然后选择 Review：

- 要根据需要手动调整 Auto Scaling 组的大小，请选择 Keep this group at its initial size (保留该组的初始大小)。有关更多信息，请参阅 [Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)。
 - 要根据您指定的条件自动调整 Auto Scaling 组的大小，选择 Use scaling policies to adjust the capacity of this group (使用扩展处理调整该组的容量)，然后按照指示操作。有关详细信息，请参阅 [配置扩展策略 \(p. 79\)](#)。
12. 在 Review 页面上，您可以选择添加标签或通知，并编辑其他配置详细信息。完成后，选择 Create Auto Scaling group (创建 Auto Scaling 组)。

标记 Auto Scaling 组和实例

通过以标签的形式将自己的元数据分配给每个组，您可以对 Auto Scaling 组进行组织和管理。可为每个标签指定键 和值。键可以是具有特定关联值的一般类别，例如“project”、“owner”或“environment”。例如，为了区分测试环境和生产环境，您可能为每个 Auto Scaling 组分配一个带有“environment”键的标签。请使用“test”值表示测试环境，或使用“production”表示生产环境。我们建议您使用一组一致的标签来帮助跟踪 Auto Scaling 组。

此外，您可以将 Auto Scaling 组中的标签传播到它启动的 Amazon EC2 实例。通过标记实例，您可以在 AWS 账单中按标签查看实例成本分配。有关更多信息，请参阅 AWS Billing and Cost Management 用户指南 中的 [使用成本分配标签](#)。

创建 Auto Scaling 组后，您可以将一个或多个标签添加到该组。您还可以添加、列出、编辑或删除现有 Auto Scaling 组的标签。

您可以控制您账户中的哪些 IAM 用户和组拥有创建、编辑或删除标签的权限。有关更多信息，请参阅 [控制对您 Amazon EC2 Auto Scaling 资源的访问权 \(p. 134\)](#)。但请记住，在实例启动后，限制您的用户对 Auto Scaling 组执行标记操作的策略不会防止阻止他们手动更改实例上的标签。有关适用于 Amazon EC2 的 IAM 策略的信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [控制对 Amazon EC2 资源的访问](#)。

Important

您也可以通过在启动模板指定标签来将标签添加到实例。但是，请务必小心，并确保您不对实例标签使用重复的键。如果这样做，Amazon EC2 Auto Scaling 将使用 Auto Scaling 组指定的相同键的值覆盖启动模板中的标签值。有关在启动模板中指定标签的信息，请参阅 [Auto Scaling 组创建启动模板 \(p. 22\)](#)。通过在启动模板中指定标签，您也可以在创建 Amazon EBS 卷时将标签添加到卷。

目录

- [标签限制 \(p. 49\)](#)
- [标记生命周期 \(p. 50\)](#)
- [对 Auto Scaling 组添加或修改标签 \(p. 50\)](#)
- [删除标签 \(p. 52\)](#)

标签限制

下面是适用于标签的基本限制：

- 每个资源的最大标签数是 50。
- 可以使用单个调用添加或删除的标签的最大数目为 25。
- 最大键长度为 128 个 Unicode 字符。
- 最大值长度为 256 个 Unicode 字符。
- 标签键和值区分大小写。
- 请勿在标签名称或值中使用 aws: 前缀，因为它预留给 AWS 使用。您不能编辑或删除具有此前缀的标签名称或值，它们不计入每个 Auto Scaling 组的标签数限制。

标记生命周期

如果您选择将标签传播到 Amazon EC2 实例，则按以下方式管理标签：

- 在大多数情况下，当 Auto Scaling 组启动实例时，它会在资源创建期间向实例添加标记，而不是在创建资源之后。
 - 例外情况是您使用启动配置启动 Spot 实例时。对于这种情况，您的 Auto Scaling 组在实例处于 Pending 生命周期状态时添加标签。如果您具有生命周期挂钩，则标签在实例进入 Pending:Wait 生命周期状态时可用。有关更多信息，请参阅 [Auto Scaling 生命周期 \(p. 6\)](#)。如果您需要 Auto Scaling 组作为启动 Spot 实例的同一 API 调用的一部分向实例添加标签，请考虑迁移到启动模板。有关更多信息，请参阅 [启动模板 \(p. 22\)](#)。
- Auto Scaling 组会自动使用 `aws:autoscaling:groupName` 键和 Auto Scaling 组的名称的值向实例添加标签。
- 您在附加现有实例时，Auto Scaling 组就会向这些实例添加标签，覆盖具有相同标签关键字的现有标签。此外，它还添加键为 `aws:autoscaling:groupName`、值为 Auto Scaling 组名称的标签。
- 将实例从 Auto Scaling 组中分离时，它仅删除 `aws:autoscaling:groupName` 标签。
- 在您手动缩减或 Auto Scaling 组自动缩减时，它会删除将要终止的实例的所有标签。

对 Auto Scaling 组添加或修改标签

当您向 Auto Scaling 组添加标签时，可以指定是否应将其添加到 Auto Scaling 组中启动的实例。如果修改标签，在更改后，标签的更新版本将添加到在 Auto Scaling 组中启动的实例。如果创建或修改 Auto Scaling 组的标签，不会对已经在 Auto Scaling 组中运行的实例进行这些更改。

目录

- [添加或修改标签 \(控制台\) \(p. 50\)](#)
- [添加或修改标签 \(AWS CLI\) \(p. 51\)](#)

添加或修改标签 (控制台)

使用 Amazon EC2 控制台：

- 创建新的 Auto Scaling 组后，将标签添加到这些组
- 添加、修改或删除现有 Auto Scaling 组的标签

创建时标记 Auto Scaling 组

当您使用 Amazon EC2 控制台创建 Auto Scaling 组时，您可以在“创建 Auto Scaling 组”向导的配置标签页上指定标签键和值。要将标签传播到在 Auto Scaling 组中启动的实例，请确保该标签的标记新实例选项保持选中状态。否则，您可以取消选择它。

添加或修改现有 Auto Scaling 组的标签

- 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
- 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
- 从列表选择一个现有的组。
- 在 Tags 选项卡上，选择 Add/Edit tags。Add/Edit Auto Scaling Group Tags (添加/编辑 Auto Scaling 组标签) 页面列出了 Auto Scaling 组的任何现有标签。
- 要修改现有标签，请编辑 Key 和 Value。
- 要添加新标签，请选择 Add tag，然后编辑 Key 和 Value。您可以使 Tag New Instances 保持选中状态，以便自动将标签添加到在 Auto Scaling 组启动的实例，否则取消选中它。

7. 添加完标签后，选择 Save。

添加或修改标签 (AWS CLI)

以下示例说明如何使用 AWS CLI 在您创建 Auto Scaling 组时添加标签，以及添加或修改现有 Auto Scaling 资源的标签。

创建时标记 Auto Scaling 组

- 使用 `create-auto-scaling-group` 命令创建一个新的 Auto Scaling 组并将标签（例如 `env=prod`）添加到 Auto Scaling 组。该标签还会添加到在 Auto Scaling 组中启动的任何实例。

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \
--launch-configuration-name my-launch-config --min-size 1 --max-size 3 \
--vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \
--tags Key=env,Value=prod,PropagateAtLaunch=true
```

创建或修改现有 Auto Scaling 组的标签

- 可以使用 `create-or-update-tags` 命令创建或修改标签。例如，以下命令将添加 `Name=my-asg` 和 `cost-center=cc123` 标签。在进行该更改后，该标签还会添加到在 Auto Scaling 组中启动的任何实例。如果具有任一键的标签已经存在，则会替换现有标签。Amazon EC2 控制台将每个实例的显示名称与为 `Name` 键指定的名称（区分大小写）关联。

```
aws autoscaling create-or-update-tags \
--tags ResourceId=my-asg,ResourceType=auto-scaling-group,Key=Name,Value=my-
asg,PropagateAtLaunch=true \
ResourceId=my-asg,ResourceType=auto-scaling-group,Key=cost-
center,Value=cc123,PropagateAtLaunch=true
```

列出 Auto Scaling 组的所有标签

- 可以使用以下 `describe-tags` 命令列出指定的 Auto Scaling 组的标签。

```
aws autoscaling describe-tags --filters Name=auto-scaling-group,Values=my-asg
```

以下为响应示例：

```
{
  "Tags": [
    {
      "ResourceType": "auto-scaling-group",
      "ResourceId": "my-asg",
      "PropagateAtLaunch": true,
      "Value": "prod",
      "Key": "env"
    }
  ]
}
```

- 或者，使用以下 `describe-auto-scaling-groups` 命令确认为 Auto Scaling 组添加了标签。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

以下为响应示例：

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupARN": "arn",
      "HealthCheckGracePeriod": 0,
      "SuspendedProcesses": [],
      "DesiredCapacity": 1,
      "Tags": [
        {
          "ResourceType": "auto-scaling-group",
          "ResourceId": "my-asg",
          "PropagateAtLaunch": true,
          "Value": "prod",
          "Key": "env"
        }
      ],
      "EnabledMetrics": [],
      "LoadBalancerNames": [],
      "AutoScalingGroupName": "my-asg",
      ...
    }
  ]
}
```

删除标签

您可以随时删除与 Auto Scaling 组关联的标签。

目录

- [删除标签 \(控制台\)](#) (p. 52)
- [删除标签 \(AWS CLI\)](#) (p. 52)

删除标签 (控制台)

删除标签

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 从列表选择一个现有的组。
4. 在 Tags 选项卡上，选择 Add/Edit tags。Add/Edit Auto Scaling Group Tags (添加/编辑 Auto Scaling 组标签) 页面列出了 Auto Scaling 组的任何现有标签。
5. 选择标签旁边的删除图标。
6. 选择 Save。

删除标签 (AWS CLI)

可以使用 `delete-tags` 命令删除标签。例如，以下命令删除键为 `env` 的标签。

```
aws autoscaling delete-tags --tags "ResourceId=my-asg,ResourceType=auto-scaling-group,Key=env"
```

您必须指定标签键，但无需指定值。如果您指定了一个值，并且该值不正确，则不会删除标签。

将负载均衡器与 Auto Scaling 组结合使用

自动扩展的目的是在需求增多时自动增大 Auto Scaling 组，并在需求减少时自动减小该组。随着容量增加或减少，所添加或删除的 Amazon EC2 实例必须在负载均衡器中注册或注销。这使得您的应用程序可以自动在这种数量不断变化的实例之间分发传入 Web 流量。

您的负载均衡器将作为到您的 Auto Scaling 组的所有传入 Web 流量的单一接触点。将实例添加到您的 Auto Scaling 组时，它需要注册到负载均衡器中，否则不会有流量路由到该实例。从 Auto Scaling 组中移除某个实例时，必须在负载均衡器中注销它，否则流量会继续路由到该实例。

要将负载均衡器与 Auto Scaling 组结合使用，请创建负载均衡器，然后将它附加到该组。

您还可以使用 Elastic Load Balancing 对实例运行状况检查以确保流量仅路由到正常实例。有关更多信息，请参阅 [向 Auto Scaling 组添加 Elastic Load Balancing 运行状况检查 \(p. 56\)](#)。

Elastic Load Balancing 类型

Elastic Load Balancing 提供了三种类型的负载均衡器可与 Auto Scaling 组配合使用：Classic Load Balancer、Application Load Balancer 和 Network Load Balancer。通过 Classic Load Balancer，实例注册到负载均衡器中。通过 Application Load Balancer 和 Network Load Balancer，实例作为一个目标注册到目标组中。

当您计划将负载均衡器与 Auto Scaling 组配合使用时，无需将 EC2 实例注册到负载均衡器或目标组。如果您启用 Elastic Load Balancing，则由 Auto Scaling 组启动的实例将自动注册到负载均衡器或目标组，并且 Auto Scaling 组终止的实例将自动从负载均衡器或目标组中注销。

传统负载均衡器

路由和负载均衡在传输层 (TCP/SSL) 或在应用程序层 (HTTP/HTTPS) 进行。传统负载均衡器支持 EC2-Classic 或 VPC。

应用程序负载均衡器

路由和负载均衡在应用程序层 (HTTP/HTTPS) 进行，并支持基于路径的路由。应用程序负载均衡器可以将请求路由到一个或多个注册目标上的端口，例如，Virtual Private Cloud (VPC) 中的 EC2 实例。

Note

应用程序负载均衡器目标组必须有一个目标类型 `instance`。有关更多信息，请参阅 Application Load Balancer 用户指南 中的 [目标类型](#)。

网络负载均衡器

路由和负载均衡在传输层 (TCP/UDP 层，即第 4 层) 进行，依据是从 TCP 数据包标头中而非从数据包内容中提取的地址信息。Network Load Balancer 可以处理突发流量、保留客户端的源 IP 以及在负载均衡器的使用寿命内使用固定 IP。

Note

网络负载均衡器目标组必须有一个目标类型 `instance`。有关更多信息，请参阅 Network Load Balancer 用户指南 中的 [目标类型](#)。

要了解有关 Elastic Load Balancing 的更多信息，请参阅以下主题：

- [什么是 Elastic Load Balancing ?](#)

- [什么是 传统负载均衡器？](#)
- [什么是 应用程序负载均衡器？](#)
- [什么是 网络负载均衡器？](#)

有关将 Amazon EC2 Auto Scaling 与 Elastic Load Balancing 集成的信息，请参阅以下主题：

主题

- [将负载均衡器附加到 Auto Scaling 组 \(p. 54\)](#)
- [向 Auto Scaling 组添加 Elastic Load Balancing 运行状况检查 \(p. 56\)](#)
- [将具有扩展和负载均衡功能的应用程序扩展到其他可用区 \(p. 57\)](#)

将负载均衡器附加到 Auto Scaling 组

Amazon EC2 Auto Scaling 与 Elastic Load Balancing 集成，以便您可以将一个或多个负载均衡器附加到现有 Auto Scaling 组。附加负载均衡器后，它自动注册组中的实例，并将传入流量分布到这些实例。

当您附加负载均衡器时，它进入 Adding 状态，同时注册组中的实例。负载均衡器注册了组中所有实例后，它进入 Added 状态。在至少一个注册实例通过运行状况检查后，它进入 InService 状态。在负载均衡器进入 InService 状态后，Amazon EC2 Auto Scaling 可以终止和替换报告为运行状况不佳的所有实例。如果注册的实例均未通过运行状况检查（例如，由于未正确配置运行状况检查），负载均衡器不会进入 InService 状态。Amazon EC2 Auto Scaling 不会终止和替换这些实例。

当分离负载均衡器时，它进入 Removing 状态，同时取消注册组中的实例。实例在取消注册后仍保持运行。如果启用连接耗尽，则 Elastic Load Balancing 将等待动态请求完成或最大超时到期（以先到者为准），然后再取消注册实例。在默认情况下，已为 Application Load Balancer 启用 Connection Draining，但它必须为 Classic Load Balancer 启用。有关更多信息，请参阅 Classic Load Balancer 用户指南 中的[连接耗尽](#)。

目录

- [先决条件 \(p. 54\)](#)
- [添加负载均衡器（控制台）\(p. 54\)](#)
- [添加负载均衡器 \(AWS CLI\) \(p. 55\)](#)

先决条件

开始前，在 Auto Scaling 组所在的 AWS 区域中创建一个负载均衡器。Elastic Load Balancing 支持三种类型的负载均衡器：Application Load Balancer、Network Load Balancer 和 Classic Load Balancer。您可以将任一类型的负载均衡器附加到 Auto Scaling 组。有关更多信息，请参阅 [Elastic Load Balancing 类型 \(p. 53\)](#)。

要配置 Auto Scaling 安全组以使用 Elastic Load Balancing 运行状况检查，请参阅[向 Auto Scaling 组添加 Elastic Load Balancing 运行状况检查 \(p. 56\)](#)。

添加负载均衡器（控制台）

使用以下过程将负载均衡器附加到现有的 Auto Scaling 组。要在创建 Auto Scaling 组时向组附加负载均衡器，请参阅 [教程：设置具有扩展和负载均衡功能的应用程序 \(p. 17\)](#)。

将负载均衡器附加到组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling Groups (Auto Scaling 组)。

3. 从列表选择一个现有的组。
4. 在 Details 选项卡上，选择 Edit。
5. 请执行下列操作之一：
 - a. [Classic Load Balancer] 对于负载均衡器，选择您的负载均衡器。
 - b. [应用程序/网络负载均衡器] 对于目标组，请选择您的目标组。
6. 选择 Save。

如果不再需要负载均衡器，请使用以下步骤将其与 Auto Scaling 组分离。

将负载均衡器与组分离

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling Groups (Auto Scaling 组)。
3. 从列表选择一个现有的组。
4. 在 Details 选项卡上，选择 Edit。
5. 请执行下列操作之一：
 - a. [Classic Load Balancer] 对于负载均衡器，删除负载均衡器。
 - b. [应用程序/网络负载均衡器] 对于目标组，请删除目标组。
6. 选择 Save。

添加负载均衡器 (AWS CLI)

附加 传统负载均衡器

可以使用以下 `attach-load-balancers` 命令将指定的负载均衡器附加到 Auto Scaling 组。

```
aws autoscaling attach-load-balancers --auto-scaling-group-name my-asg \
  --load-balancer-names my-lb
```

为应用程序负载均衡器或网络负载均衡器附加目标组

可以使用以下 `attach-load-balancer-target-groups` 命令将指定的目标组附加到 Auto Scaling 组。

```
aws autoscaling attach-load-balancer-target-groups --auto-scaling-group-name my-asg \
  --target-group-arns my-targetgroup-arn
```

分离 传统负载均衡器

如果不再需要，可以使用以下 `detach-load-balancers` 命令将负载均衡器与 Auto Scaling 组分离。

```
aws autoscaling detach-load-balancers --auto-scaling-group-name my-asg \
  --load-balancer-names my-lb
```

为应用程序负载均衡器或网络负载均衡器分离目标组

如果不再需要，可以使用以下 `detach-load-balancer-target-groups` 命令将目标组与 Auto Scaling 组分离。

```
aws autoscaling detach-load-balancer-target-groups --auto-scaling-group-name my-asg \
  --target-group-arns my-targetgroup-arn
```


向 Auto Scaling 组添加 Elastic Load Balancing 运行状况检查

Auto Scaling 组的默认运行状况检查只有 EC2 状态检查。如果一个实例未通过这些状态检查，Auto Scaling 组会认为该实例运行状况不佳并进行替换。有关更多信息，请参阅[Auto Scaling 实例的运行状况检查 \(p. 115\)](#)。

如果您已将一个或多个负载均衡器或目标组附加到 Auto Scaling 组，并且该实例未通过负载均衡器运行状况检查，则该组默认不会认为该实例运行状况不佳并进行替换。

但是，您可以选择配置该 Auto Scaling 组以使用 Elastic Load Balancing 运行状况检查。这可确保该组可以根据负载均衡器提供的其他测试确定实例的运行状况。负载均衡器会定期发送 ping、尝试进行连接或者发送请求来测试 EC2 实例。这些测试称为运行状况检查。

要了解有关 Elastic Load Balancing 运行状况检查的更多信息，请参阅以下主题：

- Classic Load Balancer 用户指南 中的[为您的 传统负载均衡器 配置运行状况检查](#)
- Application Load Balancer 用户指南 中的[目标组的运行状况检查](#)
- Network Load Balancer 用户指南 中的[目标组的运行状况检查](#)

如果您的 Auto Scaling 组配置为使用 Elastic Load Balancing 运行状况检查，而某实例未能通过 EC2 状态检查或负载均衡器运行状况检查，则它会认为该实例运行状况不佳。如果您将多个负载均衡器附加到一个 Auto Scaling 组，则只有在所有负载均衡器均报告某实例正常的情况下，它才会认为该实例正常。只要有一个负载均衡器报告某实例不正常，Auto Scaling 组就会替换该实例，即使其他负载均衡器报告该实例正常也是如此。

以下过程演示如何向 Auto Scaling 组添加 Elastic Load Balancing 运行状况检查。

目录

- [添加运行状况检查（控制台） \(p. 56\)](#)
- [添加运行状况检查 \(AWS CLI\) \(p. 56\)](#)

添加运行状况检查（控制台）

使用以下过程将宽限期为 300 秒的 ELB 运行状况检查添加到已附加负载均衡器的 Auto Scaling 组。

添加运行状况检查

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 从列表选择一个现有的组。
4. 在 Details 选项卡上，选择 Edit。
5. 对于 Health Check Type，选择 ELB。
6. 对于 Health Check Grace Period，输入 300。
7. 选择 Save。
8. 在 Instances 选项卡上，Health Status 列显示新添加的运行状况检查的结果。

添加运行状况检查 (AWS CLI)

可以使用以下 `update-auto-scaling-group` 命令创建宽限期为 300 秒的运行状况检查。

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-lb-asg \
--health-check-type ELB --health-check-grace-period 300
```

将具有扩展和负载均衡功能的应用程序扩展到其他可用区

通过使 Auto Scaling 组跨区域中的多个可用区，然后附加负载均衡器以跨这些区域分配传入流量，可以利用地理冗余的安全性和可靠性。在为负载均衡器启用的所有可用区之间均等分配传入流量。

Note

Auto Scaling 组可以包含来自同一区域中的多个可用区的 Amazon EC2 实例。不过，Auto Scaling 组不能包含来自多个区域的实例。

当一个可用区运行状况不佳或无法使用时，Amazon EC2 Auto Scaling 在不受影响的区域中启动新实例。当运行状况不佳的可用区恢复正常状态时，Amazon EC2 Auto Scaling 自动在 Auto Scaling 组的可用区中重新均匀分布实例。Amazon EC2 Auto Scaling 通过尝试在实例最少的可用区中启动新实例来执行此操作。但是，如果尝试失败，Amazon EC2 Auto Scaling 将尝试在其他可用区中启动，直到成功为止。

您可以通过将可用区添加到您的 Auto Scaling 组中，然后为您的负载均衡器启用该区域，来扩展具有扩展和负载均衡功能的应用程序的可用性。在您启用新的可用区后，负载均衡器会开始在所有已经启用的区域间均匀地路由流量。

目录

- [添加可用区 \(控制台\)](#) (p. 57)
- [添加可用区 \(AWS CLI\)](#) (p. 58)

添加可用区 (控制台)

使用以下步骤将 Auto Scaling 组扩展到其他子网 (EC2-VPC) 或可用区 (EC2-Classical)。

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling Groups (Auto Scaling 组)。
3. 从列表选择一个现有的组。
4. 在 Details 选项卡上，选择 Edit。
5. 请执行下列操作之一：
 - [EC2-VPC] 在子网中，选择与可用区相应的子网。
 - [EC2-Classical] 在可用区中，选择可用区。
6. 选择 Save。
7. 在导航窗格中，选择 Load Balancers。
8. 选择负载均衡器。
9. 请执行下列操作之一：
 - [EC2-Classical 中的 传统负载均衡器] 在 Instances (实例) 选项卡上，选择 Edit Availability Zones (编辑可用区)。在添加和删除可用区页面上，选择要添加的可用区。
 - [VPC 中的 传统负载均衡器] 在 Instances (实例) 选项卡上，选择 Edit Availability Zones (编辑可用区)。在 Add and Remove Subnets 页面上，对于 Available subnets，为要添加的子网选择加号图标 (+)。该子网将移到 Selected subnets 下。
 - [应用程序负载均衡器] 在 Description (描述) 选项卡上，为 Availability Zones (可用区) 选择 Edit (编辑)。为要添加的可用区的子网选择加号图标 (+)。该子网将移到 Selected subnets 下。

10. 选择 Save。

添加可用区 (AWS CLI)

您将使用的命令取决于您的负载均衡器是 VPC 中的 传统负载均衡器、EC2-Classic 中的 传统负载均衡器，还是 应用程序负载均衡器。

对于在 VPC 中具有 传统负载均衡器 的 Auto Scaling 组

1. 使用以下 `update-auto-scaling-group` 命令将一个子网添加到 Auto Scaling 组中。

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \
  --vpc-zone-identifier subnet-41767929 subnet-cb663da2 --min-size 2
```

2. 使用以下 `describe-auto-scaling-groups` 命令确认新子网中的实例准备好接受来自负载均衡器的流量。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

3. 使用以下 `attach-load-balancer-to-subnets` 命令为 传统负载均衡器 启用新的子网。

```
aws elb attach-load-balancer-to-subnets --load-balancer-name my-lb \
  --subnets subnet-41767929
```

对于在 EC2-Classic 中具有 传统负载均衡器 的 Auto Scaling 组

1. 使用以下 `update-auto-scaling-group` 命令将一个可用区添加到 Auto Scaling 组中。

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \
  --availability-zones us-west-2a us-west-2b us-west-2c --min-size 3
```

2. 使用以下 `describe-auto-scaling-groups` 命令确认新可用区中的实例准备好接受来自负载均衡器的流量。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

3. 使用以下 `enable-availability-zones-for-load-balancer` 命令为 传统负载均衡器 启用新的可用区。

```
aws elb enable-availability-zones-for-load-balancer --load-balancer-name my-lb \
  --availability-zones us-west-2c
```

对于具有 应用程序负载均衡器 的 Auto Scaling 组

1. 使用以下 `update-auto-scaling-group` 命令将一个子网添加到 Auto Scaling 组中。

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \
  --vpc-zone-identifier subnet-41767929 subnet-cb663da2 --min-size 2
```

2. 使用以下 `describe-auto-scaling-groups` 命令确认新子网中的实例准备好接受来自负载均衡器的流量。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

3. 使用以下 `set-subnets` 命令为 应用程序负载均衡器 启用新的子网。

```
aws elbv2 set-subnets --load-balancer-arn my-lb-arn \
```

```
--subnets subnet-41767929 subnet-cb663da2
```

在 Auto Scaling 组中启动 Spot 实例

如果您能灵活控制应用程序的运行时间并且应用程序可以中断，那么相对于按需实例，Spot 实例是经济实惠之选。本主题介绍如何通过启动配置或启动模板（而不是在 Auto Scaling 组本身）中指定设置，以便在 Auto Scaling 组中仅启动 Spot 实例。

Important

您可以将用于启动 Spot 实例的相同设置指定为 Auto Scaling 组的设置的一部分。当您将这些设置指定为 Auto Scaling 组的一部分时，您可以指定其他选项。例如，您可以指定是仅启动 Spot 实例，或者还是启动按需实例和 Spot 实例的组合。有关更多信息，请参阅 [Auto Scaling 组具有多个实例类型和购买选项](#) (p. 37)。

在使用 Amazon EC2 Auto Scaling 启动 Spot 实例之前，我们建议您熟悉使用 Amazon EC2 启动和管理 Spot 实例的操作。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [Spot 实例](#)。

在创建启动配置或启动模板以启动 Spot 实例而不是按需实例时，请注意以下事项：

- 设置您的最高价。可以将您愿意支付的最高价设置为启动配置或启动模板的一部分。
- 更改您的最高价。您必须使用新价格创建启动配置或启动模板版本。使用新启动配置时，必须将其与 Auto Scaling 组关联。在使用启动模板时，您可以将 Auto Scaling 组配置为使用默认模板或最新版本的模板。这样，它就会自动与 Auto Scaling 组相关联。只要现有实例所用的启动配置或启动模板中指定的最高价高于当前 Spot 市场价格，这些实例就会继续运行。
- Spot 市场价格与您的最高价。如果 Spot 实例的市场价格超过 Auto Scaling 组中正在运行的实例的最高价，Amazon EC2 会终止实例。如果 Spot 市场价格在最高价范围内，是否满足您的请求取决于 Spot 实例容量。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [定价和节省成本](#)。
- 维护您的 Spot 实例。Spot 实例终止后，Auto Scaling 组会尝试启动替代实例来保持该组的所需容量。如果最高价高于 Spot 市场价格，则会启动 Spot 实例。否则（或者不满足请求），它会不断尝试。
- 在可用区之间平衡。如果您指定多个可用区，Amazon EC2 Auto Scaling 会在这些可用区之间分配 Spot 请求。如果您的最高价在一个可用区中过低而无法执行任何请求，Amazon EC2 Auto Scaling 会检查请求是否已在其他区域中被执行。如果是，Amazon EC2 Auto Scaling 会取消失败的请求并在已执行请求的可用区之间重新分配它们。如果没有执行请求的可用区中的价格降得足够低以使未来的请求能够成功，Amazon EC2 Auto Scaling 会在所有可用区之间重新平衡。有关更多信息，请参阅 [再平衡活动](#) (p. 6)。
- Spot 实例终止。Amazon EC2 Auto Scaling 能够通过终止或替代按需实例的相同方式终止或替代 Spot 实例。有关更多信息，请参阅 [控制在缩小过程中终止哪些 Auto Scaling 实例](#) (p. 95)。
- Spot 中断通知。您可以使用 Spot 实例中断通知监控您的 Spot 实例的状态。例如，您可以在 Amazon CloudWatch Events 中设置一个规则，自动将 EC2 Spot 两分钟警告发送到 Amazon SNS 主题、AWS Lambda 函数或其他目标。有关更多信息，请参阅适用于 Linux 实例的 Amazon EC2 用户指南和 [Amazon CloudWatch Events 用户指南](#) 中的 [Spot 实例中断通知](#)。

将 Auto Scaling 组合并为单个多区域组

要将单独的单区域 Auto Scaling 组合并为跨多个可用区的单个组，请将其中的一个单区域组重新分区以成为多区域组。然后，删除其他组。这适用于具有或没有负载均衡器的组，但前提是新的多区域组位于与原始单区域组相同的某个可用区中。

以下示例假设您在两个不同可用区 us-west-2c 和 us-west-2a 中有两个相同的组。这两个组都具有以下规格：

- 最小大小 = 2

- 最大大小 = 5
- 所需容量 = 3

合并区域 (AWS CLI)

使用以下过程将 my-group-a 和 my-group-c 合并到一个同时涵盖 us-west-2a 和 us-west-2c 的组中。

将单独的单区域组合并为单个多区域组

1. 使用以下 [update-auto-scaling-group](#) 命令将 us-west-2c 可用区添加到 my-group-a 支持的可用区中。增加该组的最大大小以便包含两个单区域组中的实例。

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-group-a \
  --availability-zones "us-west-2a" "us-west-2c" \
  --max-size 10 --min-size 4
```

2. 使用以下 [set-desired-capacity](#) 命令增加 my-group-a 的大小。

```
aws autoscaling set-desired-capacity --auto-scaling-group-name my-group-a \
  --desired-capacity 6
```

3. (可选) 使用以下 [describe-auto-scaling-groups](#) 命令确认 my-group-a 具有新的大小。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-group-a
```

4. 使用以下 [update-auto-scaling-group](#) 命令从 my-group-c 中删除实例。

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-group-c \
  --min-size 0 --max-size 0
```

5. (可选) 使用以下 [describe-auto-scaling-groups](#) 命令确认在 my-group-c 中未剩余任何实例。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-group-c
```

下面是示例输出。

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupARN": "arn",
      "HealthCheckGracePeriod": 300,
      "SuspendedProcesses": [],
      "DesiredCapacity": 0,
      "Tags": [],
      "EnabledMetrics": [],
      "LoadBalancerNames": [],
      "AutoScalingGroupName": "my-group-c",
      "DefaultCooldown": 300,
      "MinSize": 0,
      "Instances": [],
      "MaxSize": 0,
      "VPCZoneIdentifier": "null",
      "TerminationPolicies": [
        "Default"
      ],
      "LaunchConfigurationName": "my-launch-config",
      "CreatedTime": "2015-02-26T18:24:14.449Z",
    }
  ]
}
```

```
        "AvailabilityZones": [
            "us-west-2c"
        ],
        "HealthCheckType": "EC2"
    }
}
```

6. 使用 `delete-auto-scaling-group` 命令删除 `my-group-c`。

```
aws autoscaling delete-auto-scaling-group --auto-scaling-group-name my-group-c
```

删除 Auto Scaling 基础设施

要完全删除您的扩展基础设施，请完成以下任务。

任务

- 删除 Auto Scaling 组 (p. 61)
- (可选) 删除启动配置 (p. 61)
- (可选) 删除启动模板 (p. 62)
- (可选) 删除负载均衡器 (p. 62)
- (可选) 删除 CloudWatch 警报 (p. 63)

删除 Auto Scaling 组

当您删除 Auto Scaling 组时，其所需值、最小值和最大值设置为 0。因此，将会终止实例。删除实例还会删除任何关联的日志或数据，以及该实例上的任何卷。如果不想终止一个或多个实例，您可在删除 Auto Scaling 组之前分离它们。

删除 Auto Scaling 组 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 在 Auto Scaling 组页面上，选择您的 Auto Scaling 组并依次选择操作和删除。
4. 当系统提示进行确认时，选择 Yes, Delete。

删除 Auto Scaling 组 (AWS CLI)

可以使用以下 `delete-auto-scaling-group` 命令删除 Auto Scaling 组。

```
aws autoscaling delete-auto-scaling-group --auto-scaling-group-name my-asg
```

(可选) 删除启动配置

要保留启动配置以备将来使用，可跳过此步骤。

删除启动配置 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Launch Configurations (启动配置)。

3. 在启动配置页面上，选择启动配置，然后依次选择操作和删除启动配置。
4. 当系统提示进行确认时，选择 Yes, Delete。

删除启动配置 (AWS CLI)

使用以下 `delete-launch-configuration` 命令。

```
aws autoscaling delete-launch-configuration --launch-configuration-name my-launch-config
```

(可选) 删除启动模板

您可以删除启动模板或仅删除启动模板的某个版本。在删除启动模板时，将删除其所有版本。

您可以跳过此步骤来保留启动模板以供将来使用。

删除启动模板 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择启动模板。
3. 选择启动模板，然后执行下列操作之一：
 - 选择操作，然后选择删除模板。在提示确认时，选择删除启动模板。
 - 选择操作，然后选择删除模板版本。选择要删除的版本，然后选择删除启动模板版本。

删除启动模板 (AWS CLI)

使用以下 `delete-launch-template` 命令可删除您的模板及其所有版本。

```
aws ec2 delete-launch-template --launch-template-id lt-068f72b72934aff71
```

或者，您也可以使用 `delete-launch-template-versions` 命令删除启动模板的特定版本。

```
aws ec2 delete-launch-template-versions --launch-template-id lt-068f72b72934aff71 --  
versions 1
```

(可选) 删除负载均衡器

如果未在 Elastic Load Balancing 负载均衡器中关联 Auto Scaling 组，或者要保留负载均衡器以供将来使用，请跳过该步骤。

删除您的负载均衡器 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Load Balancers。
3. 选择负载均衡器，然后依次选择操作和删除。
4. 当系统提示进行确认时，选择 Yes, Delete。

删除目标组 (控制台)

1. 在导航窗格中，选择目标组。
2. 选择目标组，然后依次选择操作、删除。

3. 当系统提示您确认时，选择 Yes。

删除与 Auto Scaling 组关联的负载均衡器 (AWS CLI)

对于 Application Load Balancer 和 Network Load Balancer，请使用以下 `delete-load-balancer` 和 `delete-target-group` 命令。

```
aws elbv2 delete-load-balancer --load-balancer-arn my-load-balancer-arn  
aws elbv2 delete-target-group --target-group-arn my-target-group-arn
```

对于 Classic Load Balancer，使用以下 `delete-load-balancer` 命令。

```
aws elb delete-load-balancer --load-balancer-name my-load-balancer
```

(可选) 删除 CloudWatch 警报

要删除与 Auto Scaling 组关联的任何 CloudWatch 警报，请完成以下步骤。

如果 Auto Scaling 组未与任何 CloudWatch 警报关联，或者您要保留警报以备将来使用，可跳过此步骤。

Note

删除 Auto Scaling 组会自动删除 Amazon EC2 Auto Scaling 为目标跟踪扩展策略管理的 CloudWatch 警报。

删除 CloudWatch 警报 (控制台)

1. 通过以下网址打开 CloudWatch 控制台：<https://console.aws.amazon.com/cloudwatch/>。
2. 在导航窗格上，选择 Alarms。
3. 选择警报，然后选择操作、删除。
4. 当系统提示进行确认时，选择 Delete。

删除 CloudWatch 警报 (AWS CLI)

使用 `delete-alarms` 命令。您可以一次删除一个或多个警报。例如，使用以下命令可删除 `Step-Scaling-AlarmHigh-AddCapacity` 和 `Step-Scaling-AlarmLow-RemoveCapacity` 警报：

```
aws cloudwatch delete-alarms --alarm-name Step-Scaling-AlarmHigh-AddCapacity Step-Scaling-AlarmLow-RemoveCapacity
```

扩展 Auto Scaling 组的大小

扩展是增加或减少应用程序的计算容量的能力。扩展从一个事件或扩展操作开始，以指示 Auto Scaling 组启动或终止 Amazon EC2 实例。

Amazon EC2 Auto Scaling 为满足您的应用程序需求提供了多种调整扩展的方式。因此，您需要很好地了解您的应用程序，这十分重要。请注意以下事项：

- Amazon EC2 Auto Scaling 应该在您的应用程序架构中发挥什么作用？通常可以将自动扩展视为一种增加和减少容量的主要方法，但自动扩展在保持稳定数量的服务器方面也是非常有用的。
- 哪些成本约束对您比较重要？由于 Amazon EC2 Auto Scaling 使用 EC2 实例，您只需为使用的资源付费。了解成本约束可以帮助您确定何时扩展应用程序以及扩展量。
- 哪些指标对应用程序比较重要？Amazon CloudWatch 支持多种可用于 Auto Scaling 组的不同指标。

目录

- [扩展选项 \(p. 64\)](#)
- [保持 Auto Scaling 组中的实例数量 \(p. 65\)](#)
- [Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)
- [适用于 Amazon EC2 Auto Scaling 的计划扩展 \(p. 74\)](#)
- [适用于 Amazon EC2 Auto Scaling 的动态扩展 \(p. 76\)](#)
- [适用于 Amazon EC2 Auto Scaling 的扩展冷却时间 \(p. 93\)](#)
- [控制在缩小过程中终止哪些 Auto Scaling 实例 \(p. 95\)](#)
- [Amazon EC2 Auto Scaling 生命周期挂钩 \(p. 100\)](#)
- [临时从 Auto Scaling 组中删除实例 \(p. 106\)](#)
- [暂停和恢复扩展流程 \(p. 110\)](#)

扩展选项

Amazon EC2 Auto Scaling 提供了多种扩展 Auto Scaling 组的方式。

始终保持当前实例级别

您可以配置 Auto Scaling 组，使其始终保持指定的运行实例数。为了保持当前实例等级，Amazon EC2 Auto Scaling 将对 Auto Scaling 组内运行的实例执行定期运行状况检查。如果 Amazon EC2 Auto Scaling 发现运行状况不佳的实例，它将终止该实例，并启动新实例。有关更多信息，请参阅[保持 Auto Scaling 组中的实例数量 \(p. 65\)](#)。

手动扩展

手动扩展是扩展资源的最基本方法，您只需指定 Auto Scaling 组的最大容量、最小容量或所需容量的变化。Amazon EC2 Auto Scaling 可以管理创建或终止实例的流程以保持更新的容量。有关更多信息，请参阅[Amazon EC2 Auto Scaling 的手动扩展 \(p. 65\)](#)。

按计划扩展

按计划扩展意味着扩展操作作为时间和日期的函数自动执行。这在您确切地知道何时增加或减少组中的实例数量时非常有用，因为该需求遵循可预测的计划。有关更多信息，请参阅[适用于 Amazon EC2 Auto Scaling 的计划扩展 \(p. 74\)](#)。

根据需求进行扩展

使用扩展策略是一种更高级的资源扩展方法，您可以定义参数以控制扩展过程。例如，您有一个当前在两个实例上运行的 Web 应用程序，并希望在应用程序负载变化时将 Auto Scaling 组的 CPU 使用率保持在 50% 左右。在根据条件变化进行扩展，但却不知道条件何时改变时，可以使用这种方法。您可以设置 Amazon EC2 Auto Scaling 以进行响应。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的动态扩展 \(p. 76\)](#)。

保持 Auto Scaling 组中的实例数量

创建 Auto Scaling 组后，Auto Scaling 组先启动足够的 EC2 实例以满足其最低容量（或其所需容量，如果指定）。如果 Auto Scaling 组没有附加其他扩展条件，Auto Scaling 组将保持这一最小数量的运行实例，即使实例运行状况不佳。

为了保持相同数量的实例，Amazon EC2 Auto Scaling 对 Auto Scaling 组内运行的实例执行定期运行状况检查。如果发现实例运行状况不佳，它将终止该实例，并启动新实例。如果您停止或终止正在运行的实例，则会将实例视为运行状况不佳并替换。有关运行状况检查替换的更多信息，请参阅 [Auto Scaling 实例的运行状况检查 \(p. 115\)](#)。

Amazon EC2 Auto Scaling 的手动扩展

您可以随时手动更改现有 Auto Scaling 组的大小。您可以更新 Auto Scaling 组的所需容量，或更新附加到 Auto Scaling 组的实例。

更改您的 Auto Scaling 组的大小（控制台）

在更改 Auto Scaling 组的大小时，Amazon EC2 Auto Scaling 将管理启动或终止实例的过程以维护新的组大小。

以下示例假设您创建了一个 Auto Scaling 组，该组的最小大小为 1，最大大小为 5。因此，该组目前正在运行的实例是 1 个。

更改 Auto Scaling 组的大小

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling Groups (Auto Scaling 组)。
3. 选择您的 Auto Scaling 组。
4. 在 Details 选项卡上，选择 Edit。
5. 对于 Desired，将所需容量增加一。例如，如果当前值是 1，请键入 2。

理想容量必须小于或等于组的最大容量。如果您的 Desired 新值大于 Max，则必须更新 Max。

完成后，选择 Save。

接下来，验证 Auto Scaling 组是否已另外启动一个实例。

验证 Auto Scaling 组的容量是否已更改

1. 在 Activity History 选项卡上，Status 列显示实例的当前状态。您可以使用刷新按钮，直至实例的状态更改为成功，这表示您的 Auto Scaling 组已成功启动新实例。
2. 在 Instances 选项卡上，Lifecycle 列显示实例的状态。启动实例只需很短的时间。实例启动后，其状态将更改为 InService。您可以看到 Auto Scaling 组启动了 1 个新实例，并且实例处于 InService 状态。

更改您的 Auto Scaling 组的大小 (AWS CLI)

在更改 Auto Scaling 组的大小时，Amazon EC2 Auto Scaling 将管理启动或终止实例的过程以维护新的组大小。默认设置是不等待默认冷却时间完成，但您可以覆盖默认行为并等待冷却时间完成。有关更多信息，请参阅[适用于 Amazon EC2 Auto Scaling 的扩展冷却时间 \(p. 93\)](#)。

以下示例假设您创建了一个 Auto Scaling 组，该组的最小大小为 1，最大大小为 5。因此，该组目前正在运行的实例是 1 个。

更改 Auto Scaling 组的大小

使用 `set-desired-capacity` 命令更改 Auto Scaling 组的大小，如以下示例所示。

```
aws autoscaling set-desired-capacity --auto-scaling-group-name my-asg \
  --desired-capacity 2
```

如果您选择遵守 Auto Scaling 组的默认冷却时间，必须指定 `--honor-cooldown` 选项，如以下示例所示。

```
aws autoscaling set-desired-capacity --auto-scaling-group-name my-asg \
  --desired-capacity 2 --honor-cooldown
```

验证 Auto Scaling 组的大小

使用 `describe-auto-scaling-groups` 命令确认 Auto Scaling 组大小已更改，如以下示例所示：

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

以下是示例输出，包含关于组和已启动实例的详细信息。

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupARN": "arn",
      "ServiceLinkedRoleARN": "arn",
      "TargetGroupARNs": [],
      "SuspendedProcesses": [],
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "Tags": [],
      "EnabledMetrics": [],
      "LoadBalancerNames": [],
      "AutoScalingGroupName": "my-asg",
      "DefaultCooldown": 300,
      "MinSize": 1,
      "Instances": [
        {
          "ProtectedFromScaleIn": false,
          "AvailabilityZone": "us-west-2a",
          "LaunchTemplate": {
            "LaunchTemplateName": "my-launch-template",
            "Version": "1",
            "LaunchTemplateId": "lt-050555ad16a3f9c7f"
          },
          "InstanceId": "i-05b4f7d5be44822a6",
          "HealthStatus": "Healthy",
          "LifecycleState": "Pending"
        }
      ]
    }
  ]
}
```

```
{
  "ProtectedFromScaleIn": false,
  "AvailabilityZone": "us-west-2a",
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "1",
    "LaunchTemplateId": "lt-050555ad16a3f9c7f"
  },
  "InstanceId": "i-0c20ac468fa3049e8",
  "HealthStatus": "Healthy",
  "LifecycleState": "InService"
},
"MaxSize": 5,
"VPCZoneIdentifier": "subnet-c87f2be0",
"HealthCheckGracePeriod": 300,
"TerminationPolicies": [
  "Default"
],
"CreatedTime": "2019-03-18T23:30:42.611Z",
"AvailabilityZones": [
  "us-west-2a"
],
"HealthCheckType": "EC2",
"NewInstancesProtectedFromScaleIn": false,
"DesiredCapacity": 2
}
]
```

注意 DesiredCapacity 显示了新值。Auto Scaling 组已启动实例。

将 EC2 实例附加到 Auto Scaling 组

Amazon EC2 Auto Scaling 为您提供了一个选项，可通过将一个或多个 EC2 实例附加到现有 Auto Scaling 组来为这些实例启用自动扩展。附加实例后，它们将成为 Auto Scaling 组的一部分。

要附加的实例必须满足以下条件：

- 实例处于 `running` 状态。
- 用于启动实例的 AMI 必须仍然存在。
- 实例不是其他 Auto Scaling 组的成员。
- 实例会启动到您的 Auto Scaling 组中定义的某个可用区。
- 如果 Auto Scaling 组具有附加的传统负载均衡器，则实例和该负载均衡器必须都位于 EC2- 或同一 VPC 中。如果 Auto Scaling 组具有附加的目标组，则实例和负载均衡器必须都位于同一 VPC 中。

当您附加实例时，该组的所需容量将增加要附加的实例数。如果要附加的实例数加上所需容量超过组的最大大小，请求将失败。

如果您将实例附加到已具有附加负载均衡器的 Auto Scaling 组，则实例将注册到该负载均衡器。如果您将实例附加到已具有附加目标组的 Auto Scaling 组，则实例将注册到该目标组。

这些示例使用的 Auto Scaling 组具有以下配置：

- Auto Scaling 组名称 = `my-asg`
- 最小大小 = 2
- 最大大小 = 5
- 所需容量 = 2

- 可用区域 = us-west-2a

附加实例（控制台）

您可以将现有实例附加到现有 Auto Scaling 组或在创建新 Auto Scaling 组时附加到新建的组。

将实例附加到新 Auto Scaling 组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择实例。
3. 选择实例。
4. 依次选择 Actions、Instance Settings 和 Attach to Auto Scaling Group。
5. 在 Attach to Auto Scaling Group 页面上，选择 a new Auto Scaling group，键入组名，然后选择 Attach。

将使用新的启动配置和为 Auto Scaling 组指定的相同名称创建新的 Auto Scaling 组。启动配置从所附加的实例获取其设置（例如，安全组和 IAM 角色）。Auto Scaling 组从所附加的实例获取设置（例如，可用区和子网），并且有所需的容量和最大大小 1。

6. （可选）要编辑 Auto Scaling 组的设置，在导航窗格中的 Auto Scaling 下选择 Auto Scaling Groups (Auto Scaling 组)。选择新 Auto Scaling 组，再选择 Edit (编辑)，根据需要更改设置，然后选择 Save (保存)。

将实例附加到现有 Auto Scaling 组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. （可选）在导航窗格上的 Auto Scaling 下，选择 Auto Scaling Groups (Auto Scaling 组)。选择 Auto Scaling 组并验证 Auto Scaling 组的最大大小足够大，可以再添加一个实例。否则，请选择 Edit，增加最大大小，然后选择 Save。
3. 在导航窗格中，选择实例。
4. 选择实例。
5. 依次选择 Actions、Instance Settings 和 Attach to Auto Scaling Group。
6. 在 Attach to Auto Scaling Group 页面上，选择 an existing Auto Scaling group，选择实例，然后选择 Attach。
7. 如果实例不符合条件，则会显示一条错误消息并提供详细信息。例如，实例可能没有位于与 Auto Scaling 组相同的可用区中。选择 Close 并使用符合条件的实例重试。

附加实例 (AWS CLI)

将实例附加到 Auto Scaling 组

1. 描述使用以下 `describe-auto-scaling-groups` 命令特定的 Auto Scaling 组。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg
```

以下示例响应显示所需的容量为 2，并且组具有两个运行的实例：

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupARN": "arn",
      "ServiceLinkedRoleARN": "arn",
```

```
{
  "TargetGroupARNs": [],
  "SuspendedProcesses": [],
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "1",
    "LaunchTemplateId": "lt-050555ad16a3f9c7f"
  },
  "Tags": [],
  "EnabledMetrics": [],
  "LoadBalancerNames": [],
  "AutoScalingGroupName": "my-asg",
  "DefaultCooldown": 300,
  "MinSize": 1,
  "Instances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "HealthStatus": "Healthy",
      "LifecycleState": "Pending"
    },
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-0c20ac468fa3049e8",
      "HealthStatus": "Healthy",
      "LifecycleState": "InService"
    }
  ],
  "MaxSize": 5,
  "VPCZoneIdentifier": "subnet-c87f2be0",
  "HealthCheckGracePeriod": 300,
  "TerminationPolicies": [
    "Default"
  ],
  "CreatedTime": "2019-03-18T23:30:42.611Z",
  "AvailabilityZones": [
    "us-west-2a"
  ],
  "HealthCheckType": "EC2",
  "NewInstancesProtectedFromScaleIn": false,
  "DesiredCapacity": 2
}
]
```

2. 使用以下 `attach-instances` 命令将实例附加到 Auto Scaling 组。

```
aws autoscaling attach-instances --instance-ids i-0787762faf1c28619 --auto-scaling-group-name my-asg
```

3. 要验证是否已附加实例，请使用 `describe-auto-scaling-groups` 命令。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg
```


下面的示例响应显示所需容量已从 1 增加到 3，并且有一个新实例 i-0787762faf1c28619：

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupARN": "arn",
      "ServiceLinkedRoleARN": "arn",
      "TargetGroupARNs": [],
      "SuspendedProcesses": [],
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "Tags": [],
      "EnabledMetrics": [],
      "LoadBalancerNames": [],
      "AutoScalingGroupName": "my-asg",
      "DefaultCooldown": 300,
      "MinSize": 1,
      "Instances": [
        {
          "ProtectedFromScaleIn": false,
          "AvailabilityZone": "us-west-2a",
          "LaunchTemplate": {
            "LaunchTemplateName": "my-launch-template",
            "Version": "1",
            "LaunchTemplateId": "lt-050555ad16a3f9c7f"
          },
          "InstanceId": "i-05b4f7d5be44822a6",
          "HealthStatus": "Healthy",
          "LifecycleState": "Pending"
        },
        {
          "ProtectedFromScaleIn": false,
          "AvailabilityZone": "us-west-2a",
          "LaunchTemplate": {
            "LaunchTemplateName": "my-launch-template",
            "Version": "1",
            "LaunchTemplateId": "lt-050555ad16a3f9c7f"
          },
          "InstanceId": "i-0c20ac468fa3049e8",
          "HealthStatus": "Healthy",
          "LifecycleState": "InService"
        },
        {
          "ProtectedFromScaleIn": false,
          "AvailabilityZone": "us-west-2a",
          "LaunchTemplate": {
            "LaunchTemplateName": "my-launch-template",
            "Version": "1",
            "LaunchTemplateId": "lt-050555ad16a3f9c7f"
          },
          "InstanceId": "i-0787762faf1c28619",
          "HealthStatus": "Healthy",
          "LifecycleState": "InService"
        }
      ],
      "MaxSize": 5,
      "VPCZoneIdentifier": "subnet-c87f2be0",
      "HealthCheckGracePeriod": 300,
      "TerminationPolicies": [
        "Default"
      ]
    }
  ]
}
```

```
        "CreatedTime": "2019-03-18T23:30:42.611Z",
        "AvailabilityZones": [
            "us-west-2a"
        ],
        "HealthCheckType": "EC2",
        "NewInstancesProtectedFromScaleIn": false,
        "DesiredCapacity": 3
    }
}
```

从 Auto Scaling 组中分离 EC2 实例

可以从 Auto Scaling 组中删除实例。分离实例之后，可以独立于 Auto Scaling 组的其余部分管理这些实例。通过分离实例，您可以：

- 将实例从一个 Auto Scaling 组移出，然后将其附加到其他组。有关更多信息，请参阅[将 EC2 实例附加到 Auto Scaling 组 \(p. 67\)](#)。
- 通过使用正在运行应用程序的现有实例创建 Auto Scaling 组，以测试该组，然后在测试完成后将这些实例与 Auto Scaling 组分离。

分离实例时，您可以选择按分离的实例数减少 Auto Scaling 组的所需容量。如果您选择不减少容量，Amazon EC2 Auto Scaling 会启动新实例以替换分离的实例。如果减少容量，但将多个实例从同一可用区分离，则除非暂停 AZRebalance 流程，否则 Amazon EC2 Auto Scaling 可以重新平衡可用区。有关更多信息，请参阅[扩展流程 \(p. 110\)](#)。

如果要分离的实例数将 Auto Scaling 组的大小降到最小容量以下，您必须先减少该组的最小容量，然后才能分离实例。

如果将实例从具有附加负载均衡器的 Auto Scaling 组分离，则该实例将从负载均衡器取消注册。如果您从具有附加目标组的 Auto Scaling 组中分离实例，则实例将从目标组取消注册。如果为负载均衡器启用了连接耗尽，则 Amazon EC2 Auto Scaling 会等待未送达请求完成。

这些示例使用的 Auto Scaling 组具有以下配置：

- Auto Scaling 组名称 = my-asg
- 最小容量 = 1
- 最大容量 = 5
- 所需容量 = 4
- 可用区域 = us-west-2a

分离实例 (控制台)

使用以下步骤将实例与 Auto Scaling 组分离。

将实例与现有 Auto Scaling 组分离

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Instances 选项卡上，选择实例，然后选择 Actions 和 Detach。
5. 在分离实例页面上，选中该复选框可启动替换实例，取消选中该复选框可减少所需容量。选择分离实例。

分离实例 (AWS CLI)

使用以下步骤将实例与 Auto Scaling 组分离。

将实例与现有 Auto Scaling 组分离

1. 使用以下 [describe-auto-scaling-instances](#) 命令列出当前实例。

```
aws autoscaling describe-auto-scaling-instances
```

以下示例响应显示该组有 4 个运行的实例：

```
{
  "AutoScalingInstances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "InService"
    },
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-0c20ac468fa3049e8",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "InService"
    },
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-0787762faf1c28619",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "InService"
    },
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-0f280a4c58d319a8a",
      "AutoScalingGroupName": "my-asg",

```

```
        "HealthStatus": "HEALTHY",  
        "LifecycleState": "InService"  
      }  
    ]  
  }  
}
```

2. 使用以下 [detach-instances](#) 命令分离实例并减少所需的容量。

```
aws autoscaling detach-instances --instance-ids i-05b4f7d5be44822a6 \  
  --auto-scaling-group-name my-asg --should-decrement-desired-capacity
```

3. 使用以下 [describe-auto-scaling-instances](#) 命令验证是否已分离该实例。

```
aws autoscaling describe-auto-scaling-instances
```

以下示例响应显示现在有 3 个运行的实例：

```
{  
  "AutoScalingInstances": [  
    {  
      "ProtectedFromScaleIn": false,  
      "AvailabilityZone": "us-west-2a",  
      "LaunchTemplate": {  
        "LaunchTemplateName": "my-launch-template",  
        "Version": "1",  
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"  
      },  
      "InstanceId": "i-0c20ac468fa3049e8",  
      "AutoScalingGroupName": "my-asg",  
      "HealthStatus": "HEALTHY",  
      "LifecycleState": "InService"  
    },  
    {  
      "ProtectedFromScaleIn": false,  
      "AvailabilityZone": "us-west-2a",  
      "LaunchTemplate": {  
        "LaunchTemplateName": "my-launch-template",  
        "Version": "1",  
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"  
      },  
      "InstanceId": "i-0787762faf1c28619",  
      "AutoScalingGroupName": "my-asg",  
      "HealthStatus": "HEALTHY",  
      "LifecycleState": "InService"  
    },  
    {  
      "ProtectedFromScaleIn": false,  
      "AvailabilityZone": "us-west-2a",  
      "LaunchTemplate": {  
        "LaunchTemplateName": "my-launch-template",  
        "Version": "1",  
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"  
      },  
      "InstanceId": "i-0f280a4c58d319a8a",  
      "AutoScalingGroupName": "my-asg",  
      "HealthStatus": "HEALTHY",  
      "LifecycleState": "InService"  
    }  
  ]  
}
```

适用于 Amazon EC2 Auto Scaling 的计划扩展

按计划扩展使您可以按照可预测的负载变化来设置您自己的扩展计划。例如，您的 Web 应用程序的流量会在每周的星期三开始增加，并在星期四保持高流量状态，然后在星期五开始下降。您可以根据 Web 应用程序的可预测流量模式来计划扩展操作。扩展操作会作为时间和日期的函数自动执行。

Note

要根据可预测的负载变化进行扩展，您还可以使用 AWS Auto Scaling 的预测扩展功能。有关更多信息，请参阅 [AWS Auto Scaling 用户指南](#)。

要配置 Auto Scaling 组以根据计划进行扩展，您需要创建一个计划的操作。计划的操作指示 Amazon EC2 Auto Scaling 在指定的时间执行扩展操作。要创建计划的扩展操作，请指定扩展操作应生效的开始时间，以及用于扩展操作的新的最小大小、最大大小和所需大小。在指定的时间，Amazon EC2 Auto Scaling 将用扩展操作指定的最小大小、最大大小和所需大小值来更新组。

您可以创建仅用于一次扩展的预先计划操作，或者创建用于按经常性计划进行扩展的预先计划操作。

注意事项

创建计划的操作时，请记住以下内容。

- 可以保证同组内计划操作的执行顺序正确，但不保证跨组的计划操作的执行顺序正确。
- 计划操作的执行时间一般为几秒钟。然而，该操作可能会比计划的开始时间延迟最多两分钟。因为系统将按照指定操作的顺序来执行 Auto Scaling 组内的操作，所以计划开始时间彼此接近的计划操作可能需要更长时间才能执行。
- 您最多可以为每个 Auto Scaling 组创建 125 个计划的操作。
- 计划的操作必须具有唯一时间值。如果您尝试计划在某个时间进行活动，但另一个扩展活动已计划占用该时间，系统将拒绝该请求，并向您发送关于该冲突的出错消息。
- 在到达计划操作的结束时间后，它不会保留在您的账户中。
- 您可以暂时禁用计划扩展而不删除您的计划操作。有关更多信息，请参阅[暂停和恢复扩展流程 \(p. 110\)](#)。
- 不支持冷却时间。
- 您也可以为 Amazon EC2 以外的资源计划扩展操作。有关更多信息，请参阅 Application Auto Scaling 用户指南中的[计划扩展](#)。

创建和管理计划操作（控制台）

您可以使用控制台创建仅扩展一次或按重复计划扩展的计划操作。完成以下过程可创建计划的操作以扩展您的 Auto Scaling 组。

为 Auto Scaling 组创建计划操作

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Scheduled Actions 选项卡上，选择 Create Scheduled Action。
5. 在 Create Scheduled Action 页面上，执行以下操作：
 - 使用至少一个下列值来指定组大小：Min、Max 或 Desired Capacity。
 - 为 Recurrence 选择一个选项。如果您选择一次，则在指定的时间执行操作。如果您选择 Cron，请输入一个 Cron 表达式，此表达式指定执行操作的时间（采用 UTC 时间）。如果您选择以 Every 开头的选项，则将为您创建 Cron 表达式。
 - 如果您为 Recurrence 选择了 Once，则在 Start Time 中指定操作的时间。

- 如果您指定了重复计划，则可以为 Start Time 和 End Time 指定值。如果您指定开始时间，则这是最早执行操作的时间。如果您指定结束时间，在该时间后将不执行操作。

6. 选择 Create。

更新计划的操作

如果您的需求发生变化，则可更新计划的操作。

更新计划的操作

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Scheduled Actions 选项卡上，选择计划的操作。
5. 选择 Actions 和 Edit。
6. 在 Edit Scheduled Action 页面上，执行以下操作：
 - 使用 Min、Max 或 Desired Capacity 根据需要更新组的大小。
 - 根据需要更新指定的重复。
 - 根据需要更新开始时间和结束时间。
 - 选择保存。

删除计划的操作

当您不再需要计划操作时，可以将其删除。

删除计划操作

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Scheduled Actions 选项卡上，选择计划的操作。
5. 依次选择 Actions 和 Delete。
6. 当系统提示进行确认时，选择是，删除。

创建和管理计划操作 (AWS CLI)

您可以使用 `put-scheduled-update-group-action` 命令，创建并更新仅扩展一次或按重复计划扩展的已计划的操作。

仅扩展一次

您可以指定一次性计划，在特定日期和时间（采用 UTC 时间）自动扩展您的 Auto Scaling 组。

- 要在特定时间减少 Auto Scaling 组中正在运行的实例数量，请使用以下命令。在为 `--start-time` 指定的日期和时间，如果组中当前包含多个实例，则该组缩减到 1 个实例。

```
aws autoscaling put-scheduled-update-group-action --scheduled-action-name my-one-time-action \
--auto-scaling-group-name my-asg --start-time "2019-05-13T08:00:00Z" --desired-capacity 1
```

- 要在特定时间增加 Auto Scaling 组中正在运行的实例数量，请使用以下命令。在为 `--start-time` 指定的日期和时间，如果组中当前已少于 3 个实例，则该组会扩展到 3 个实例。

```
aws autoscaling put-scheduled-update-group-action --scheduled-action-name my-one-time-action \
  --auto-scaling-group-name my-asg --start-time "2019-05-12T08:00:00Z" --desired-capacity 3
```

按照重复计划扩展

您可以使用 Unix cron 语法格式指定重复的时间表（采用 UTC 时间）。此格式包含五个空格分隔的字段：[Minute] [Hour] [Day_of_Month] [Month_of_Year] [Day_of_Week]。有关此格式的更多信息，请参阅 [Crontab](#)。

使用以下 `put-scheduled-update-group-action` 命令创建一个计划操作，该操作在每年的 1 月、6 月和 12 月的第一天的 00:30 运行。

```
aws autoscaling put-scheduled-update-group-action --scheduled-action-name my-recurring-action \
  --auto-scaling-group-name my-asg --recurrence "30 0 1 1,6,12 *" --desired-capacity 3
```

删除计划的操作

删除计划操作

使用以下 `delete-scheduled-action` 命令。

```
aws autoscaling delete-scheduled-action --scheduled-action-name my-recurring-action
```

适用于 Amazon EC2 Auto Scaling 的动态扩展

当您配置动态扩展时，必须定义如何根据不断变化的需求进行扩展。例如，您有一个当前在两个实例上运行的 Web 应用程序，并希望在应用程序负载变化时将 Auto Scaling 组的 CPU 使用率保持在 50% 左右。这为您提供额外容量以处理流量高峰，而无需维护过多的空闲资源。您可以将 Auto Scaling 组配置为自动扩展来满足这一需求。策略类型决定扩展操作如何执行。

目录

- [扩展策略类型 \(p. 76\)](#)
- [多个扩展策略 \(p. 77\)](#)
- [适用于 Amazon EC2 Auto Scaling 的目标跟踪扩展策略 \(p. 77\)](#)
- [适用于 Amazon EC2 Auto Scaling 的简单扩展和分步扩展策略 \(p. 81\)](#)
- [向现有的 Auto Scaling 组添加扩展策略 \(p. 88\)](#)
- [基于 Amazon SQS 进行扩展 \(p. 89\)](#)
- [删除扩展策略 \(p. 92\)](#)

扩展策略类型

Amazon EC2 Auto Scaling 支持以下类型的扩展策略：

- **目标跟踪扩展** — 根据特定指标的目标值增加或减少组的当前容量。它的工作方式类似于您家中负责控制温度的恒温器 - 您只需选择一个温度，恒温器会替您完成所有其他工作。

- 分步扩展 — 根据一组扩展调整，增加或减小组的当前容量，这些调整称为分步调整，将根据警报严重程度发生变化。
- 简单扩展 — 根据一个扩展调整值增加或减小当前组容量。

如果您根据一个按 Auto Scaling 组中的实例数成比例增减的使用率指标进行扩展，我们建议您使用目标跟踪扩展策略。否则，我们建议您使用步进扩展策略。

多个扩展策略

在大多数情况下，目标跟踪扩展策略就足以将您的 Auto Scaling 组配置为自动扩展或缩减。目标跟踪扩展策略允许您选择所需结果，并让 Auto Scaling 组根据需要添加和删除实例以实现该结果。

对于高级扩展配置，您的 Auto Scaling 组可以有多个扩展策略。例如，您可以定义一个或多个目标跟踪扩展策略，一个或多个步进扩展策略，或者同时使用两种策略。这样可以更灵活地覆盖多种场景。

为了说明多个策略如何协同工作，请设想一个应用程序，它使用一个 Auto Scaling 组和 Amazon SQS 队列将请求发送到单个 EC2 实例。为了帮助确保应用程序性能达到最佳级别，有两个策略用于控制何时扩展 Auto Scaling 组。一个是使用自定义指标、根据队列中的 SQS 消息数增加和移除容量的目标跟踪策略。另一个是步进策略，它使用 Amazon CloudWatch CPUUtilization 指标、当实例在指定时间长度内的使用率超过 90% 时增加容量。

在同时实施多个策略时，各个策略可能会同时指示 Auto Scaling 组扩展（或缩减）。例如，在 EC2 实例触发 CPUUtilization 指标的 CloudWatch 警报的同时，SQS 队列可能会触发自定义指标的警报。

如果发生上述情况，Amazon EC2 Auto Scaling 会选择在扩展和缩减时都提供最大容量的策略。例如，假定 CPU 使用率策略启动一个实例，而 SQS 队列的策略启动两个实例。如果同时满足两个策略的扩展条件，Amazon EC2 Auto Scaling 优先选择 SQS 队列策略。这会导致 Auto Scaling 组启动两个实例。

即使策略使用不同的扩展条件，使提供最大容量的策略具有优先级的方法也适用。例如，如果一个策略终止三个实例，另一个策略将实例数量减少 25%，且在扩展时组中有八个实例，则 Amazon EC2 Auto Scaling 会优先考虑为组提供最大数量实例的策略。这会导致 Auto Scaling 组终止两个实例（ $8 \times 25\% = 2$ ）。目的是防止 Amazon EC2 Auto Scaling 删除过多实例。

适用于 Amazon EC2 Auto Scaling 的目标跟踪扩展策略

在使用目标跟踪扩展策略时，可以选择一个扩展指标并设置一个目标值。Amazon EC2 Auto Scaling 创建和管理触发扩展策略的 CloudWatch 警报，并根据指标和目标值计算扩展调整。扩展策略根据需要增加或减少容量，将指标保持在指定的目标值或接近指定的目标值。除了将指标保持在目标值附近以外，目标跟踪扩展策略还会对由于负载模式变化而造成的指标变化进行调整。

例如，您可以使用目标跟踪扩展进行以下操作：

- 配置目标跟踪扩展策略，使 Auto Scaling 组的平均聚合 CPU 利用率保持在 40%。
- 为 Auto Scaling 组配置目标跟踪扩展策略，使 Elastic Load Balancing 目标组的每目标请求数保持在 1000。

建议将随 Amazon EC2 实例指标扩展的频率设置为 1 分钟，这可更快地响应使用率变化。如果将随指标扩展的频率设置为 5 分钟，可能会导致响应速度变慢，并根据过期的指标数据进行扩展。默认情况下，Amazon EC2 实例是针对基本监控启用的，也就是说，实例的指标数据以 5 分钟的间隔提供。您可以启用详细监控，从而以 1 分钟的频率获取实例的指标数据。有关更多信息，请参阅 [为 Auto Scaling 实例配置监控 \(p. 119\)](#)。

注意事项

请注意以下事项：

- 目标跟踪扩展策略假设它应该在指定指标高于目标值时扩展 Auto Scaling 组。因此，不能使用目标跟踪扩展策略在指定指标低于目标值时扩展 Auto Scaling 组。
- 您可能会看到目标值与实际指标数据点之间存在差距。这是因为我们在确定需要添加或删除多少个实例时通过向上或向下取整来保守地进行操作，以免添加的实例数量不足或删除的实例数量过多。但是，对于实例较少的 Auto Scaling 组，组的使用率可能偏离目标值较远。例如，您将 CPU 使用率的目标值设置为 50%，然后 Auto Scaling 组超过了该目标。我们可以确定，添加 1.5 个实例会将 CPU 使用率降低到接近 50%。由于不可能添加 1.5 个实例，我们将该值向上取整，添加两个实例。这可能会将 CPU 使用率降至 50% 以下，但可确保应用程序具有充足的支持资源。同样，如果我们确定删除 1.5 个实例可使 CPU 使用率提高到 50% 以上，我们将只删除一个实例。
- 对于包含更多实例的更大 Auto Scaling 组，使用率分布在更多实例上，在这种情况下，添加或删除实例会导致目标值与实际指标数据点之间的差距缩小。
- 为了确保应用程序可用性，Auto Scaling 组会针对指标尽快按比例向外扩展，但会逐渐向内扩展。
- 一个 Auto Scaling 组可以同时实施多个扩展策略。有关更多信息，请参阅 [多个扩展策略 \(p. 77\)](#)。
- 您可以为 Auto Scaling 组提供多个目标跟踪扩展策略，前提是它们各自使用不同的指标。Amazon EC2 Auto Scaling 的目的是始终优先考虑可用性，因此其行为会有所不同，具体取决于目标跟踪策略是否已准备好扩展或缩减。如果任何目标跟踪策略已准备好进行扩展，它将扩展 Auto Scaling 组，但仅在所有目标跟踪策略（启用了缩减部分）准备好缩减时才执行缩减。
- 您可以禁用目标跟踪扩展策略的缩减部分。通过使用该功能，您可以灵活地使用不同的方法缩减 Auto Scaling 组。例如，您可以使用不同的扩展策略类型进行缩减，同时使用目标跟踪扩展策略进行扩展。
- 不要编辑或删除为目标跟踪扩展策略配置的 CloudWatch 警报。当您删除扩展策略时，将自动删除与目标跟踪扩展策略关联的 CloudWatch 警报。
- 您还可以为 Amazon EC2 以外的资源使用目标跟踪扩展。有关更多信息，请参阅 Application Auto Scaling 用户指南 中的 [目标跟踪扩展策略](#)。

选择指标

您可以使用 Amazon EC2 控制台根据预定义的指标应用目标跟踪扩展策略。另外，您可以使用 Amazon EC2 Auto Scaling CLI 或 API 以根据预定义或自定义的指标来应用扩展策略。可供使用的预定义指标如下：

- ASGAverageCPUUtilization—Auto Scaling 组的平均 CPU 使用率。
- ASGAverageNetworkIn—Auto Scaling 组在所有网络接口上收到的平均字节数。
- ASGAverageNetworkOut—Auto Scaling 组在所有网络接口上发送的平均字节数。
- ALBRequestCountPerTarget—Application Load Balancer 目标组中每个目标完成的请求数。

您可以通过指定自定义指标来选择其他可用 Amazon CloudWatch 指标。

选择指标时请记住原则：

- 并非所有指标都适用于目标跟踪。当指定自定义指标时，这可能很重要。指标必须是有效的使用率指标，它用于描述实例的繁忙程度。指标值必须随着 Auto Scaling 组中的实例数按比例增加或减少。这样，指标数据可用于随实例数量按比例扩展或缩减。例如，如果某个 Auto Scaling 组的负载分布在各个实例上，则该 Auto Scaling 组的 CPU 使用率指标（即指标维度为 AutoScalingGroupName 的 Amazon EC2 指标 CPUUtilization）能够正常工作。
- 以下指标不适用于目标跟踪：
 - Auto Scaling 组前面的负载均衡器收到的请求数（即，Elastic Load Balancing 指标 RequestCount）。负载均衡器收到的请求数不会根据 Auto Scaling 组的使用率而发生变化。
 - 负载均衡器请求延迟（即，Elastic Load Balancing 指标 Latency）。请求延迟可能会根据使用率增加而增加，但不一定按比例变化。
 - CloudWatch SQS 队列指标 ApproximateNumberOfMessagesVisible。队列中的消息数可能不会随着处理队列中的消息的 Auto Scaling 组的大小按比例发生变化。但是，用于测量消息数（Auto Scaling 组的每个 EC2 实例的队列中）的自定义指标能够正常工作。有关更多信息，请参阅 [基于 Amazon SQS 进行扩展 \(p. 89\)](#)。

- 当指定指标数据不足时，目标跟踪扩展策略不会扩展 Auto Scaling 组，除非您使用 `ALBRequestCountPerTarget` 指标。其中的原理是 `ALBRequestCountPerTarget` 指标为不存在关联数据的时间段发出零，而且目标跟踪处理需要指标数据来解释低使用率趋势。要实现在不存在路由到目标组的请求时使您的 Auto Scaling 组缩减为 0 个实例，组的最小容量必须设置为 0。

创建具有目标跟踪扩展策略的 Auto Scaling 组（控制台）

创建具有目标跟踪扩展策略的 Auto Scaling 组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择 Create Auto Scaling group。
4. 在 Create Auto Scaling Group 页面上，执行以下操作之一：
 - 选择 Create an Auto Scaling group from an existing launch configuration，选择现有的启动配置，然后选择 Next Step (下一步)。
 - 如果您没有想要使用的启动配置，请选择 Create a new launch configuration，然后按照指示操作。有关更多信息，请参阅 [创建启动配置 \(p. 28\)](#)。
5. 在 Configure Auto Scaling group details 页面上，执行以下操作：
 - a. 对于 Group name (组名)，键入 Auto Scaling 组的名称。
 - b. 对于 Group size (组大小)，键入 Auto Scaling 组所需的容量。
 - c. 如果启动配置指定需要 VPC 的实例（如 T2 实例），则必须从 Network 选择一个 VPC。另外，如果您的 AWS 账户支持 EC2-Classic 并且这些实例不需要 VPC，您可以选择 Launch into EC2-Classic 或 VPC。
 - d. 如果您在上一步中选择了 VPC，请从 Subnet 中选择一个或多个子网。如果您在上一步中已选择 EC2-Classic，请从 Availability Zone(s) 中选择一个或多个可用区。
 - e. 选择 Next: Configure scaling policies。
6. 在 Configure scaling policies (配置扩展策略) 页面上，执行以下操作：
 - a. 选择 Use scaling policies to adjust the capacity of this group (使用扩展策略调整此组的容量)。
 - b. 使用以 Scale between (扩展范围) 开头的行指定 Auto Scaling 组的最小大小和最大大小。例如，如果组已经是最大大小，则指定一个新的最大值才能进行扩展。
 - c. 对于 Scale Group Size，指定扩展策略。您可以选择为该策略指定名称，然后为指标类型选择一个值。
 - d. 为指标指定 Target value。
 - e. 为实例需要指定实例预热值，用于控制新启动的实例在多长时间后开始作用于 CloudWatch 指标。
 - f. 选中 Disable scale-in 选项则只创建扩展策略。这样，可以根据需要创建独立的其他类型的缩减策略。
 - g. 选择 Review。
 - h. 在 Review 页面上，选择 Create Auto Scaling group。

实例预热

您可以指定新启动实例的预热时间（秒）。在指定预热时间过期前，实例不会计入 Auto Scaling 组的聚合指标。

在扩展时，我们不会将正在预热的实例看作当前组容量的一部分。这可确保添加的实例不会超出您的需要。

在缩减时，我们会将正在终止的实例视为当前组容量的一部分。因此不会从 Auto Scaling 组中删除更多实例。

正在进行扩展活动时不能启动收缩活动。

创建目标跟踪扩展策略 (AWS CLI)

按如下方式使用 AWS CLI 为 Auto Scaling 组配置目标跟踪扩展策略。

任务

- [步骤 1：创建 Auto Scaling 组 \(p. 80\)](#)
- [步骤 2：创建目标跟踪扩展策略 \(p. 80\)](#)

步骤 1：创建 Auto Scaling 组

可以使用 `create-auto-scaling-group` 命令通过启动配置 `my-launch-config` 创建一个名为 `my-asg` 的 Auto Scaling 组。如果您没有要使用的启动配置，可以创建一个。有关更多信息，请参阅 [create-launch-configuration](#)。

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
--launch-configuration-name my-launch-config \  
--vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  
--max-size 5 --min-size 1
```

步骤 2：创建目标跟踪扩展策略

您可以创建目标跟踪扩展策略，指示 Auto Scaling 组在应用程序负载变化时动态地增加或减少组中正在运行的 EC2 实例数量。

示例：目标跟踪配置文件

下面是一个目标跟踪配置文件示例 (应将其另存为 `config.json`)：

```
{  
  "TargetValue": 40.0,  
  "PredefinedMetricSpecification":  
    {  
      "PredefinedMetricType": "ASGAverageCPUUtilization"  
    }  
}
```

或者，您可以通过创建自定义指标规范并为 CloudWatch 中的每个参数添加值来自定义用于扩展的指标。以下是一个示例目标跟踪配置，它将指定指标的平均利用率保持在 40%。

```
{  
  "TargetValue":40.0,  
  "CustomizedMetricSpecification":{  
    "MetricName":"MyUtilizationMetric",  
    "Namespace":"MyNamespace",  
    "Dimensions":[  
      {  
        "Name":"MyOptionalMetricDimensionName",  
        "Value":"MyOptionalMetricDimensionValue"  
      }  
    ],  
    "Statistic":"Average",  
    "Unit":"Percent"  
  }  
}
```

示例：cpu40-target-tracking-scaling-policy

使用 `put-scaling-policy` 命令以及之前创建的 `config.json` 文件创建一个名为 `cpu40-target-tracking-scaling-policy` 的扩展策略，它将 Auto Scaling 组的平均 CPU 使用率保持在 40%：

```
aws autoscaling put-scaling-policy --policy-name cpu40-target-tracking-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \
  --target-tracking-configuration file://config.json
```

适用于 Amazon EC2 Auto Scaling 的简单扩展和分步扩展策略

借助简单和步进扩展策略，您可以为触发扩展流程的 CloudWatch 警报选择扩展指标和阈值，并定义当阈值违反指定数量的评估期时应如何扩展您的 Auto Scaling 组。

即使您只进行单次扩展调整，我们也建议您使用步进扩展策略而不是简单扩展策略。Amazon EC2 Auto Scaling 起初只支持简单扩展策略。如果您在引入目标跟踪和步进策略前创建了自己的扩展策略，您的策略将被视为简单扩展策略。有关更多信息，请参阅 [简单扩展策略和冷却时间](#) (p. 81)。

如果您的扩展指标是一个随 Auto Scaling 组中实例数成比例增减的使用率指标，我们建议您改用目标跟踪扩展策略。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的目标跟踪扩展策略](#) (p. 77)。您仍然可以选择使用目标跟踪扩展与步进扩展来实现更高级的扩展策略配置。例如，如果需要，您可以在利用率达到特定级别时配置更积极的响应。

步进扩展策略

步进扩展策略根据一组扩展调整增加或减少 Auto Scaling 组的当前容量，这些调整称为步进调整。这些调整将根据警报违约规模发生变化。

扩展活动启动后，该策略继续响应其他警报，甚至在进行扩展活动或健康检查替换时也是如此。因此，Amazon EC2 Auto Scaling 会在接收警报消息时评估任何突破的警报。

如果要创建扩展策略，可以指定影响聚合指标的新启动实例就绪所需的估计预热时间。有关更多信息，请参阅 [实例预热](#) (p. 83)。

您还可以为 Amazon EC2 以外的资源使用步进扩展。有关更多信息，请参阅 Application Auto Scaling 用户指南中的 [步进扩展策略](#)。

简单扩展策略和冷却时间

大多数情况下，分步扩展策略优于简单扩展策略。对于简单扩展，在启动扩展活动后，该策略必须等待扩展活动或运行状况检查替换完成并且冷却时间到期，然后才会响应其他警报。冷却时间有助于防止在先前活动产生明显影响前启动其他扩展活动。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的扩展冷却时间](#) (p. 93)。

Amazon EC2 Auto Scaling 不支持对分步扩展策略指定冷却时间。因此，对此类策略不能指定冷却时间，组的默认冷却时间也不适用。

扩展调整类型

执行步进扩展或简单扩展策略时，它将使用策略中指定的扩展调整来更改 Auto Scaling 组的当前容量。扩展调整不能将组容量更改为超出最大组大小或低于最小组大小。

对于步进扩展和简单扩展，Amazon EC2 Auto Scaling 支持以下调整类型：

- `ChangeInCapacity` — 将当前组容量增加或减少指定的实例数。正值将增加容量，负调整值将减小容量。

示例：如果当前组容量为 3 个实例，调整值为 5，则当此策略执行时，会向组中添加 5 个实例，总共达到 8 个实例。

- **ExactCapacity** — 将当前组容量更改为指定的实例数。为该调整类型指定一个正值。

示例：如果当前组容量为 3 个实例，调整值为 5，则当此策略执行时，容量将设置为 5 个实例。

- **PercentChangeInCapacity** — 将当前组容量增加或减少指定的百分比。正值将增加容量，负值将减小容量。如果得出的值不是整数，将按如下所示进行取整：
 - 大于 1 的值向下取整。例如，12.7 取整为 12。
 - 0 和 1 之间的值舍入到 1。例如，.67 取整为 1。
 - 0 和 -1 之间的值舍入到 -1。例如，-.58 取整为 -1。
 - 小于 -1 的值向上取整。例如，-6.67 取整为 -6。

示例：如果当前容量为 10 个实例，调整值为 10%，则当此策略执行时，将向组中添加 1 个实例，总共达到 11 个实例。

对于 **PercentChangeInCapacity**，您还可以指定要扩展的最小实例数（使用 **MinAdjustmentMagnitude** 参数或控制台中的添加实例时至少需添加）。例如，假定您创建一个增加 25% 的策略，并且您指定最小增量为 2 个实例。如果您有一个 Auto Scaling 组包含 4 个实例，而要执行该扩展策略，则 4 的 25% 就是 1 个实例。但是，因为您指定了最小增量 2，则将添加 2 个实例。

分步调整

创建分步扩展策略时，将添加一个或多个分步调整，让您可以根据警报的严重程度进行扩展。每个分步调整指定以下内容：

- 指标值的下限
- 指标值的上限
- 要扩展的数量（基于扩展调整类型）

策略分步调整有一些规则：

- 分步调整范围不能重叠或有间隙。
- 只有一个分步调整可以有空下限（负无穷）。如果一个分步调整有负下限，则必须有一个分步调整有空下限。
- 只有一个分步调整可以有空上限（正无穷）。如果一个分步调整有正上限，则必须有一个分步调整有空上限。
- 同一分步调整中的上限和下限不能为空。
- 如果指标值高于违例阈值，则含下限而不含上限。如果指标值低于违例阈值，则不含下限而含上限。

CloudWatch 根据与您的 CloudWatch 警报关联的指标的统计信息聚合指标数据点。当突破警报阈值时，将触发相应的扩展策略。Amazon EC2 Auto Scaling 将聚合类型应用于 CloudWatch 中最新的指标数据点（与原始指标数据相对）。它将此聚合指标值与步进调整定义的上限和下限进行比较，以确定执行哪个步进调整。

如果您使用 AWS 管理控制台，则以绝对值的形式指定上限和下限。如果您使用 API 或 CLI，则以违例阈值相对值的形式指定上限和下限。例如，假设您有一个警报的违例阈值为 50，扩展调整类型为 **PercentChangeInCapacity**。还有具有以下分步调整的扩展和缩减策略：

Scale out policy			
Lower bound	Upper bound	Adjustment	Metric value

0	10	0	50 <= 值 < 60
10	20	10	60 <= 值 < 70
20	null	30	70 <= 值 < +无穷
Scale in policy			
Lower bound	Upper bound	Adjustment	Metric value
-10	0	0	40 < 值 <= 50
-20	-10	-10	30 < 值 <= 40
null	-20	-30	- 无穷 < 值 <= 30

您的组的当前容量和所需容量均为 10 个实例。当聚合指标值大于 40 小于 60 时，该组维持其当前容量和所需容量。

如果指标值达到 60，根据扩展策略的第二个分步调整 (增加 10 个实例的 10%)，组的所需容量将增加 1 个实例，总共达到 11 个实例。在新实例运行并且其指定预热时间过期后，当前组容量将增加到 11 个实例。如果在容量增加后指标值又增加到 70，根据扩展策略的第三个分步调整 (增加 11 个实例的 30%，即 3.3 个实例，向下取整为 3 个实例)，组的所需容量将增加另外 3 个实例，总共达到 14 个实例。

如果指标值降到 40，根据缩减策略的第二个分步调整 (删除 14 个实例的 10%，向下取整为 1 个实例)，组的所需容量将减小 1 个实例，达到 13 个实例。如果在容量减小后指标值又降到 30，根据缩减策略的第三个分步调整 (删除 13 个实例的 30%，即 3.9 个实例，向下取整为 3 个实例)，组的所需容量将减少另外 3 个实例，达到 10 个实例。

实例预热

通过分步扩展策略可以指定新启动实例的预热时间（秒数）。在指定预热时间过期前，实例不会计入 Auto Scaling 组的聚合指标。在扩展时，AWS 也不会将正在预热的实例看作当前组容量的一部分。因此，落入同一分步调整中的多个警报违例只会导致一个扩展活动。这可确保添加的实例不会超出您的需要。

使用上一部分中的示例，假设指标值增加到 60，而在新实例仍在预热时又增加到 62。当前容量仍为 10 个实例，因此，添加了 1 个实例（10 个实例的 10%）。不过，组的所需容量已经是 11 个实例，因此，AWS 不会进一步增加所需的容量。如果指标在新实例仍在预热时增长到 70，我们应添加 3 个实例（10 个实例的 30%）。但该组的所需容量已经是 11，所以我们只添加 2 个实例，因为新的所需容量为 13 个实例。

在缩减时，AWS 将正在终止的实例视为当前组容量的一部分。因此不会从 Auto Scaling 组中删除更多实例。

正在进行扩展活动时不能启动收缩活动。

创建带有分步扩展策略的 Auto Scaling 组（控制台）

您可以创建一个使用 CloudWatch 警报来确定 Auto Scaling 组何时应缩减或扩展的扩展策略。每个 CloudWatch 警报均监控一个指标，并在该指标超过您在策略中指定的阈值时向 Amazon EC2 Auto Scaling 发送消息。您可以使用警报来监控从您使用的 AWS 中的服务发送到 CloudWatch 的任何指标。或者，您可以创建和监控自己的指标。

创建 CloudWatch 警报后，可以指定 Amazon SNS 主题，以在警报状态改变时发送电子邮件通知。有关更多信息，请参阅[创建 Amazon CloudWatch 警报 \(p. 121\)](#)。

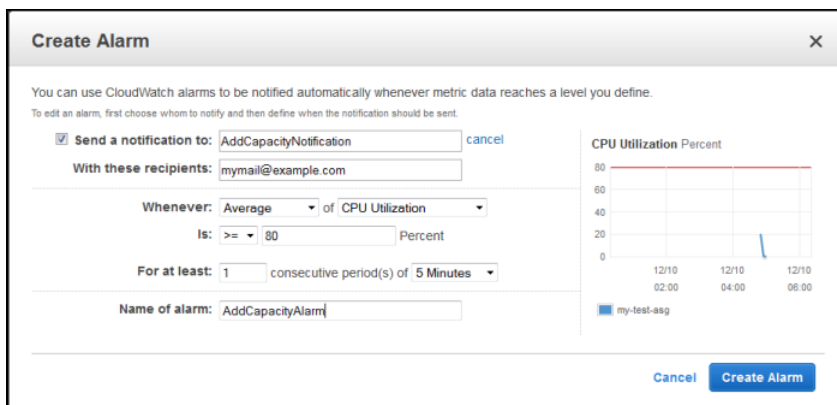
使用控制台可以创建具有两个扩展策略的 Auto Scaling 组：将组容量增加 30% 的扩展策略，以及将组容量减小到两个实例的收缩策略。

创建带有分步扩展策略的 Auto Scaling 组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择 Create Auto Scaling group。
4. 在 Create Auto Scaling Group 页面上，执行以下操作之一：
 - 选择 Create an Auto Scaling group from an existing launch configuration，选择现有的启动配置，然后选择 Next Step (下一步)。
 - 如果您没有想要使用的启动配置，请选择 Create a new launch configuration，然后按照指示操作。有关更多信息，请参阅[创建启动配置 \(p. 28\)](#)。
5. 在 Configure Auto Scaling group details 页面上，执行以下操作：
 - a. 对于 Group name (组名)，键入 Auto Scaling 组的名称。
 - b. 对于 Group size (组大小)，键入 Auto Scaling 组所需的容量。
 - c. 如果启动配置指定需要 VPC 的实例（如 T2 实例），则必须从 Network 选择一个 VPC。另外，如果您的 AWS 账户支持 EC2-Classic 并且这些实例不需要 VPC，您可以选择 Launch into EC2-Classic 或 VPC。
 - d. 如果您在上一步中选择了 VPC，请从 Subnet 中选择一个或多个子网。如果您在上一步中已选择 EC2-Classic，请从 Availability Zone(s) 中选择一个或多个可用区。
 - e. 选择 Next: Configure scaling policies。
6. 在 Configure scaling policies (配置扩展策略) 页面上，执行以下操作：
 - a. 选择 Use scaling policies to adjust the capacity of this group (使用扩展策略调整此组的容量)。
 - b. 使用以 Scale between (扩展范围) 开头的行指定 Auto Scaling 组的最小大小和最大大小。例如，如果组已经是最大大小，则需要指定一个新的最大值才能进行扩展。

Scale between and instances. These will be the minimum and maximum size of your group.

- c. 在 Increase Group Size 下指定扩展策略。您可以选择为该策略指定名称，然后选择 Add new alarm。
- d. 在 Create Alarm 页面上，选择 create topic。对于 Send a notification to，键入 SNS 主题的名称。对于 With these recipients，键入一个或多个用于接收通知的电子邮件地址。可以使用自定义名称替换默认警报名称。接下来，为策略指定指标和标准。例如，您可以保留 Whenever 的默认设置 (CPU 使用率平均值)。对于 Is，选择 \geq 并键入 80%。对于 For at least，键入 1 个连续的 5 Minutes 时间段。选择 Create Alarm。



- e. 对于 Take the action，选择 Add，在下一个字段中键入 30，然后选择 percent of group。默认情况下，此步骤调整的下限为警报阈值，上限为 null (正无穷)。

要添加其他步骤调整，请选择 Add step。要设置可扩展的最少实例数，请更新 Add instances in increments of at least 1 instance(s) 中的数量字段。

(可选) 建议您采用默认值通过几个步骤创建这些扩展策略。要创建简单扩展策略, 请选择 [Create a simple scaling policy](#)。有关更多信息, 请参阅[适用于 Amazon EC2 Auto Scaling 的简单扩展和分步扩展策略 \(p. 81\)](#)。

Increase Group Size

Name: AddCapacity

Execute policy when: AddCapacityAlarm [Edit](#) [Remove](#)
breaches the alarm threshold: CPUUtilization >= 80 for 300 seconds
for the metric dimensions AutoScalingGroupName = my-asg

Take the action: Add 30 percent of group when 80 <= CPUUtilization < +infinity
[Add step](#) ⓘ

Add instances in increments of at least 7 instance(s)

Instances need: 300 seconds to warm up after each step

[Create a simple scaling policy](#) ⓘ

- f. 为 Instances need (实例需要) 指定实例预热值, 用于控制新启动的实例在多长时间后开始作用于 CloudWatch 指标。
- g. 在 Decrease Group Size 下指定缩减策略。您可以选择为该策略指定名称, 然后选择 Add new alarm。
- h. 在 Create Alarm 页面上, 您可以选择为扩展策略创建的通知, 也可以为缩减策略创建一个新通知。可以使用自定义名称替换默认警报名称。保留 Whenever 的默认设置 (CPU 使用率平均值)。对于 Is, 选择 <= 并键入 40%。对于 For at least, 键入 1 个连续的 5 Minutes 时间段。选择 Create Alarm。
- i. 对于 Take the action, 选择 Remove, 在下一个字段中键入 2, 然后选择 instances。默认情况下, 此步骤调整的上限为警报阈值, 下限为 null (负无穷)。要添加其他步骤调整, 请选择 Add step。

(可选) 建议您采用默认值通过几个步骤创建这些扩展策略。要创建简单扩展策略, 请选择 [Create a simple scaling policy](#)。有关更多信息, 请参阅[扩展策略类型 \(p. 76\)](#)。

Decrease Group Size

Name: DecreaseCapacity

Execute policy when: DecreaseCapacityAlarm [Edit](#) [Remove](#)
breaches the alarm threshold: CPUUtilization <= 40 for 300 seconds
for the metric dimensions AutoScalingGroupName = my-asg

Take the action: Remove 2 instances when 40 >= CPUUtilization > -infinity
[Add step](#) ⓘ

[Create a simple scaling policy](#) ⓘ

- j. 选择 Review。
 - k. 在 Review 页面上, 选择 Create Auto Scaling group。
7. 使用以下步骤验证您的 Auto Scaling 组的扩展策略。
- a. Auto Scaling Group creation status 页面确认您的 Auto Scaling 组已成功创建。选择 View your Auto Scaling Groups (查看您的 Auto Scaling 组)。
 - b. 在 Auto Scaling Groups 页面上, 选择您刚创建的 Auto Scaling 组。
 - c. 在 Activity History (活动历史记录) 选项卡上, Status (状态) 列显示 Auto Scaling 组是否具有已成功启动的实例。
 - d. 在 Instances 选项卡上, Lifecycle 列包含实例的状态。启动实例只需很短的时间。实例启动后, 其生命周期状态更改为 InService。

Health Status (运行状态) 列显示对您的实例进行 EC2 实例运行状况检查的结果。

- e. 在扩展策略选项卡上，可以查看为 Auto Scaling 组创建的策略。

配置分步扩展策略 (AWS CLI)

按如下方式使用 AWS CLI 为 Auto Scaling 组配置分步扩展策略。

任务

- [步骤 1：创建 Auto Scaling 组 \(p. 86\)](#)
- [第 2 步：创建扩展策略 \(p. 86\)](#)
- [第 3 步：创建 CloudWatch 警报 \(p. 87\)](#)

步骤 1：创建 Auto Scaling 组

使用以下 [create-auto-scaling-group](#) 命令，通过使用启动配置 `my-launch-config` 创建一个名为 `my-asg` 的 Auto Scaling 组。如果您没有要使用的启动配置，可以创建一个。有关更多信息，请参阅 [create-launch-configuration](#)。

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \
  --launch-configuration-name my-launch-config \
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \
  --max-size 5 --min-size 1
```

第 2 步：创建扩展策略

您可以创建分步或简单扩展策略，指示 Auto Scaling 在应用程序负载发生变化时该怎么做。

示例：my-step-scaleout-policy

使用以下 [put-scaling-policy](#) 命令创建名为 `my-step-scaleout-policy` 的分步扩展策略，其调整类型为 `PercentChangeInCapacity`、基于以下分步增加组容量（假设 CloudWatch 警报阈值为 60%）：

- 当指标值大于或等于 70% 但小于 80% 时，将实例计数增加 10%
- 当指标值大于或等于 80% 但小于 90% 时，将实例计数增加 20%
- 当指标值大于或等于 90% 时，将实例计数增加 30%

```
aws autoscaling put-scaling-policy \
  --auto-scaling-group-name my-asg \
  --policy-name my-step-scaleout-policy \
  --policy-type StepScaling \
  --adjustment-type PercentChangeInCapacity \
  --metric-aggregation-type Average \
  --step-adjustments
  MetricIntervalLowerBound=10.0,MetricIntervalUpperBound=20.0,ScalingAdjustment=10 \
  MetricIntervalLowerBound=20.0,MetricIntervalUpperBound=30.0,ScalingAdjustment=20 \
  MetricIntervalLowerBound=30.0,ScalingAdjustment=30 \
  --min-adjustment-magnitude 1
```

输出包括策略的 ARN。将此 ARN 保存在安全位置。您需要使用它来创建 CloudWatch 警报。

```
{
  "PolicyARN": "arn:aws:autoscaling:region:123456789012:scalingPolicy:4ee9e543-86b5-4121-b53b-aa4c23b5bbcc:autoScalingGroupName/my-asg:policyName/my-step-scalein-policy"
}
```

```
}
```

示例：my-step-scalein-policy

使用以下 `put-scaling-policy` 命令创建名为 my-step-scalein-policy 的分步扩展策略，其调整类型为 `ChangeInCapacity`、将组容量减少 2 个实例：

```
aws autoscaling put-scaling-policy \
  --auto-scaling-group-name my-asg \
  --policy-name my-step-scalein-policy \
  --policy-type StepScaling \
  --adjustment-type ChangeInCapacity \
  --step-adjustments MetricIntervalUpperBound=0.0,ScalingAdjustment=-2
```

输出包括作为策略唯一名称的 ARN。随后，您可以使用 ARN 或者策略名称和组名称的组合来指定策略。将此 ARN 保存在安全位置。您需要使用它来创建 CloudWatch 警报。

```
{
  "PolicyARN": "arn:aws:autoscaling:region:123456789012:scalingPolicy:ac542982-
cbeb-4294-891c-a5a941dfa787:autoScalingGroupName/my-asg:policyName/my-step-scaleout-policy"
}
```

简单扩展策略示例

或者，您可以使用以下 CLI 命令而不是前文中的 CLI 命令创建简单扩展策略。这些命令的输出包括每个策略的 ARN，您在创建 CloudWatch 警报时需要它。请记住，使用简单扩展策略会取代冷却时间。

示例：my-simple-scaleout-policy

使用以下 `put-scaling-policy` 命令创建名为 my-simple-scaleout-policy 的简单扩展策略，其调整类型为 `PercentChangeInCapacity`、将组容量增加 30%：

```
aws autoscaling put-scaling-policy --policy-name my-simple-scaleout-policy \
  --auto-scaling-group-name my-asg --scaling-adjustment 30 \
  --adjustment-type PercentChangeInCapacity
```

示例：my-simple-scalein-policy

使用以下 `put-scaling-policy` 命令创建名为 my-simple-scalein-policy 的简单扩展策略，其调整类型为 `ChangeInCapacity`、将组容量减少两个实例：

```
aws autoscaling put-scaling-policy --policy-name my-simple-scalein-policy \
  --auto-scaling-group-name my-asg --scaling-adjustment -2 \
  --adjustment-type ChangeInCapacity
```

第 3 步：创建 CloudWatch 警报

在步骤 2 中，您已创建扩展策略，这些策略可以指示 Auto Scaling 组在指定条件发生变化时如何缩减和扩展。在此步骤中，您将通过标识要监控的指标、定义用于扩展的条件来创建警报，然后将警报与扩展策略相关联。

示例：AddCapacity

使用以下 CloudWatch `put-metric-alarm` 命令创建警报，基于至少两个连续评估期（两分钟）内 60% 的平均 CPU 阈值增加 Auto Scaling 组的大小。要使用您自己的自定义指标，请在 `--metric-name` 中指定其名称，并在 `--namespace` 指定其命名空间。

```
aws cloudwatch put-metric-alarm --alarm-name Step-Scaling-AlarmHigh-AddCapacity \
--metric-name CPUUtilization --namespace AWS/EC2 --statistic Average \
--period 120 --evaluation-periods 2 --threshold 60 \
--comparison-operator GreaterThanOrEqualToThreshold \
--dimensions "Name=AutoScalingGroupName,Value=my-asg" \
--alarm-actions PolicyARN
```

示例：RemoveCapacity

使用以下 CloudWatch `put-metric-alarm` 命令创建警报，基于至少两个连续评估期（两分钟）内 40% 的平均 CPU 阈值减小 Auto Scaling 组的大小。要使用您自己的自定义指标，请在 `--metric-name` 中指定其名称，并在 `--namespace` 指定其命名空间。

```
aws cloudwatch put-metric-alarm --alarm-name Step-Scaling-AlarmLow-RemoveCapacity \
--metric-name CPUUtilization --namespace AWS/EC2 --statistic Average \
--period 120 --evaluation-periods 2 --threshold 40 \
--comparison-operator LessThanOrEqualToThreshold \
--dimensions "Name=AutoScalingGroupName,Value=my-asg" \
--alarm-actions PolicyARN
```

向现有的 Auto Scaling 组添加扩展策略

使用控制台将扩展策略添加到现有的 Auto Scaling 组。扩展策略用于动态增加或减少组中正在运行的 EC2 实例的数量，以符合不断变化的条件。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的动态扩展 \(p. 76\)](#)。

根据指标更新具有扩展的 Auto Scaling 组

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择 Auto Scaling 组。
4. 在 Scaling Policies 选项卡上，选择 Add policy。

页面会显示所有可用资源。

5. 如果要添加目标跟踪扩展策略，请按照以下子步骤操作。如果使用的是简单或步进扩展策略，请跳到下一步。
 - a. 在 Name 中键入策略的名称。
 - b. 选择 Metric type 并为该指标指定 Target value。
 - c. 为 Instances need (实例需要) 指定实例预热值，用于控制新启动的实例在多长时间后开始作用于 CloudWatch 指标。
 - d. 如果只想创建扩展策略，请选中 Disable scale-in 选项。这样，可以根据需要创建独立的缩减策略。
 - e. 选择 Create。
6. 如果您使用分步扩展策略，请选择创建一个带有步骤的扩展策略，然后执行以下操作：

Note

建议您采用选项通过几个步骤创建这些扩展策略。要使用简单的扩展策略，请改为选择 Create a simple scaling policy。有关简单和步进扩展的更多信息，请参阅[适用于 Amazon EC2 Auto Scaling 的简单扩展和分步扩展策略 \(p. 81\)](#)。

- a. 对于 Name，键入策略的名称，然后选择 Create new alarm。
- b. 在 Create Alarm 页面上，选择 create topic。对于 Send a notification to，键入 SNS 主题的名称。对于 With these recipients，键入一个或多个用于接收通知的电子邮件地址。可以使用自定义名称

替换默认警报名称。下一步，使用 Whenever、Is 和 For at least 为该警报指定指标和标准。选择 Create Alarm。

- c. 使用 Take the action 为策略指定扩展活动。默认情况下，此步骤调整的下限为警报阈值，上限为 null（正无穷）。要添加其他步骤调整，请选择 Add step。
- d. 为 Instances need (实例需要) 指定实例预热值，用于控制新启动的实例在多长时间后开始作用于 CloudWatch 指标。
- e. 选择 Create。

基于 Amazon SQS 进行扩展

本部分向您介绍如何根据 Amazon Simple Queue Service (Amazon SQS) 队列的需求变化来扩展 Auto Scaling 组。Amazon SQS 提供了安全、持久且可用的托管队列，可让您集成和分离分布式软件系统和组件。如果您不熟悉 Amazon SQS，请参阅 [Amazon Simple Queue Service 开发人员指南](#) 以了解更多信息。

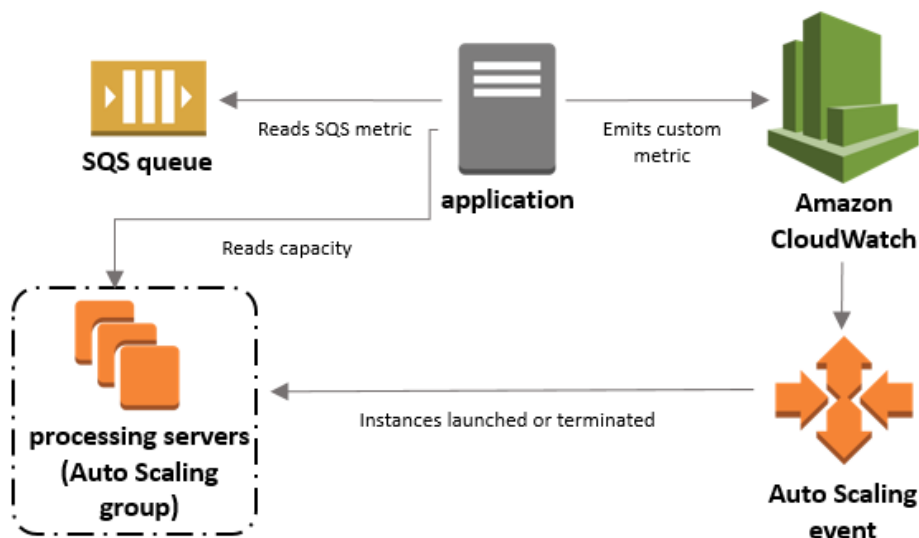
在多种情况下，您可能会考虑根据 Amazon SQS 队列中的活动扩展。假设您有 Web 应用程序，让用户上传图像并联机使用。每个图像需要先调整大小并编码，然后才能发布。该应用程序在 Auto Scaling 组中的 EC2 实例上运行，已配置为处理您的典型上传速率。不正常的实例将终止并进行替换，以始终保持当前的实例等级。该应用程序将图像的原始位图数据放在 Amazon SQS 队列中以进行处理。它处理这些图像，然后将处理的图像发布到某个位置以供用户查看。

如果上传的图像数不随时间波动，此架构可以良好运作。如果上传级别发生变化，会发生什么情况？如果上传按照可预见的计划增长和降低，您可以指定执行扩展活动的时间和日期。有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的计划扩展 \(p. 74\)](#)。按策略扩展是一种更为动态的 Auto Scaling 组扩展方法，可让您定义参数以控制扩展过程。例如，您可以创建策略，用于在平均上传数达到特定水平时扩展 EC2 实例的队列。在根据条件变化进行扩展，但却不知道条件何时改变时，可以使用这种方法。

此配置有三个主要部分：

- 一个 Auto Scaling 组，用于管理处理来自 SQS 队列的消息的 EC2 实例。
- 一个发送到 Amazon CloudWatch 的自定义指标，用于衡量 Auto Scaling 组中每个 EC2 实例的队列的消息数。
- 一个目标跟踪策略，配置您的 Auto Scaling 组根据自定义指标和设定的目标值进行扩展。CloudWatch 警报调用该扩展策略。

下图演示了此配置的架构。



选择有效的指标和目标值

Amazon SQS 队列中的消息数不仅仅定义所需的实例数。实际上，队列中的实例数可能由多种因素决定，包括处理消息所需的时间以及可接受的延迟长度（队列延迟）。

该解决方案使用要维护的每个实例的积压指标以及每个实例的可接受积压的目标值。您可以按以下所示计算这些数字：

- 每个实例的积压：要确定您每个实例的积压，首先通过 Amazon SQS 指标 `ApproximateNumberOfMessages` 确定 SQS 队列的长度（可从队列中检索的消息数）。将该数字除以队列的运行容量（对于 Auto Scaling 组，这是处于 `InService` 状态的实例数量），以获得每个实例的积压。
- 每个实例的可接受积压：要确定您的目标值，请先计算您的应用程序可以接受的延迟。然后，将可接受的延迟值除以 EC2 实例处理一条消息所用的平均时间。

我们来举例说明，当前 `ApproximateNumberOfMessages` 为 1500，队列的运行容量为 10。如果每条消息的平均处理时间为 0.1 秒，最长可接受延迟为 10 秒，则每个实例的可接受积压为 $10/0.1$ ，即 100。这意味着您的目标跟踪策略的目标值为 100。由于当前每个实例的积压为 $(1500/10)$ ，您的队列增加 5 个实例以维持与目标值的比例。

以下示例创建自定义指标和目标跟踪扩展策略，根据这些计算来配置您的 Auto Scaling 组进行扩展。

配置 Amazon SQS 扩展

以下部分向您说明如何使用 AWS CLI 为 SQS 队列设置自动扩展。前面的过程假设您已经有一个队列（标准队列或 FIFO 队列）、一个 Auto Scaling 组以及 EC2 实例（运行使用队列的应用程序）。

任务

- [步骤 1：创建 CloudWatch 自定义指标](#) (p. 90)
- [步骤 2：创建目标跟踪扩展策略](#) (p. 91)
- [步骤 3：测试扩展策略](#) (p. 92)

步骤 1：创建 CloudWatch 自定义指标

首先，从您的 AWS 账户中读取指标以创建自定义计算指标。然后，计算前面的章节中建议的每个实例的积压指标。最后，按照 1 分钟的粒度将该数字发布到 CloudWatch。

只要可能，您应该按照 1 分钟的频率根据 EC2 实例指标进行扩展（也称为详细监控），因为这会确保更快地响应使用率变化。如果将随指标扩展的频率设置为 5 分钟，可能会导致响应速度变慢，并根据过期的指标数据进行扩展。默认情况下，为 EC2 实例启用基本监控，也就是说，实例的指标数据以 5 分钟为间隔提供。您可以启用详细监控，以按照 1 分钟的频率获取实例的指标数据。有关更多信息，请参阅 [使用 Amazon CloudWatch 监控 Auto Scaling 组和实例](#) (p. 117)。

创建 CloudWatch 自定义指标

1. 使用 SQS `get-queue-attributes` 命令获取在队列中等待的消息数 (`ApproximateNumberOfMessages`)：

```
aws sqs get-queue-attributes --queue-url https://sqs.region.amazonaws.com/123456789/MyQueue \
    --attribute-names ApproximateNumberOfMessages
```

2. 使用 `describe-auto-scaling-groups` 命令获取组的运行容量，这是处于 `InService` 生命周期状态的实例数。此命令返回 Auto Scaling 组的实例及其生命周期状态。


```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg
```

3. 通过将 ApproximateNumberOfMessages 指标除以队列的运行容量指标来计算每个实例的积压。
4. 使用 AWS CLI 或 API，按照 1 分钟的粒度将结果发布为 CloudWatch 自定义指标。自定义指标是使用您选择的指标名称和命名空间定义的。自定义指标的命名空间不能以“AWS/”开头。有关发布自定义指标的更多信息，请参阅 Amazon CloudWatch 用户指南 中的 [发布自定义指标](#) 主题。

以下是一个 CLI `put-metric-data` 命令的示例。

```
aws cloudwatch put-metric-data --metric-name MyBacklogPerInstance --  
namespace MyNamespace \  
--unit None --value 20 --  
dimensions MyOptionalMetricDimensionName=MyOptionalMetricDimensionValue
```

应用程序发出所需的指标之后，数据发送到 CloudWatch。该指标会显示在 CloudWatch 控制台中。您可以登录到 AWS 管理控制台并导航到 CloudWatch 页面来访问它。然后，通过导航到指标页面或者使用搜索框搜索指标来查看指标。如需有关查看指标方面的帮助，请参阅 Amazon CloudWatch 用户指南 中的 [查看可用指标](#)。

步骤 2：创建目标跟踪扩展策略

然后，创建目标跟踪扩展策略，指示 Auto Scaling 组在应用程序负载变化时动态地增加或减少组中正在运行的 EC2 实例。您可以使用目标跟踪扩展策略，因为扩展指标是与组容量成比例增加或减少的使用率指标。

创建目标跟踪扩展策略

1. 使用以下命令可以在您的主目录的 `config.json` 文件中为扩展策略指定目标值。

对于 `TargetValue`，计算每个实例的可接受积压指标并在此处输入。要计算此数值，请考虑正常延迟值并将其除以处理一条消息所需的平均时间。

```
$ cat ~/config.json  
{  
  "TargetValue":100,  
  "CustomizedMetricSpecification":{  
    "MetricName":"MyBacklogPerInstance",  
    "Namespace":"MyNamespace",  
    "Dimensions":[  
      {  
        "Name":"MyOptionalMetricDimensionName",  
        "Value":"MyOptionalMetricDimensionValue"  
      }  
    ],  
    "Statistic":"Average",  
    "Unit":"None"  
  }  
}
```

2. 使用 `put-scaling-policy` 命令以及在前面的步骤中创建的 `config.json` 文件创建扩展策略：

```
aws autoscaling put-scaling-policy --policy-name my-scaling-policy --auto-scaling-  
group-name my-asg \  
--policy-type TargetTrackingScaling --target-tracking-configuration file:///~/  
config.json
```

这会创建两个警报：一个用于增加实例数量，另一个用于减少实例数量。它还将返回注册到 CloudWatch 的策略的 Amazon 资源名称 (ARN)，CloudWatch 使用该名称在突破指标时调用扩展。

步骤 3：测试扩展策略

在您完成设置后，验证您的扩展策略是否正常工作。要进行测试，可以增加 SQS 队列中的消息数，然后验证您的 Auto Scaling 组是否启动另一个 EC2 实例，也可以减少 SQS 队列中的消息数，然后验证 Auto Scaling 组是否终止一个 EC2 实例。

测试扩展函数

1. 按照[教程：将消息发送到 Amazon SQS 队列](#)中的步骤将消息添加到队列。请确保您增加了队列中的消息数量，使得每个实例的积压指标超过目标值。

您的更改触发 CloudWatch 警报可能需要几分钟时间。

2. 使用 `describe-auto-scaling-groups` 命令确认该组已启动一个实例：

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

测试缩减函数

1. 按照[教程：将消息发送到 Amazon SQS 队列](#)中的步骤从队列中删除消息。请确保您减少了队列中的消息数量，使得每个实例的积压指标低于目标值。

您的更改触发 CloudWatch 警报可能需要几分钟时间。

2. 使用 `describe-auto-scaling-groups` 命令确认该组已终止一个实例：

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

删除扩展策略

当您不再需要某个扩展策略时，可将其删除。您可能还需要删除 CloudWatch 警报。删除目标跟踪扩展策略还将删除任何关联的 CloudWatch 警报。删除步进扩展策略或简单扩展策略会删除底层警报操作，但不删除与扩展策略关联的 CloudWatch 警报，即使它不再具有关联的操作。

删除扩展策略（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择 Auto Scaling 组。
4. 在扩展策略选项卡上，选择操作、删除。
5. 当系统提示进行确认时，选择 Yes, Delete。
6. （可选）如果您删除步进扩展策略或简单扩展策略，请执行以下操作来删除与策略关联的 CloudWatch 警报。您可以跳过此步骤来保留该警报以备将来使用。
 - a. 通过以下网址打开 CloudWatch 控制台：<https://console.aws.amazon.com/cloudwatch/>。
 - b. 在导航窗格上，选择 Alarms。
 - c. 选择警报（例如 `awsec2-my-auto-scaling-group-CPU-Utilization`），然后选择操作、删除。
 - d. 当系统提示进行确认时，选择 Delete。

删除扩展策略 (AWS CLI)

使用以下 `delete-policy` 命令：

```
aws autoscaling delete-policy --auto-scaling-group-name my-asg \  
--policy-name my-step-scalein-policy
```

删除您的 CloudWatch 警报 (AWS CLI)

对于步进扩展策略和简单扩展策略，请使用 [delete-alarms](#) 命令删除与策略关联的 CloudWatch 警报。您可以跳过此步骤来保留该警报以备将来使用。您可以一次删除一个或多个警报。例如，使用以下命令可删除 Step-Scaling-AlarmHigh-AddCapacity 和 Step-Scaling-AlarmLow-RemoveCapacity 警报：

```
aws cloudwatch delete-alarms --alarm-name Step-Scaling-AlarmHigh-AddCapacity Step-Scaling-  
AlarmLow-RemoveCapacity
```

适用于 Amazon EC2 Auto Scaling 的扩展冷却时间

冷却时间有助于确保在以前的扩展活动生效之前，Auto Scaling 组不会启动或终止其他实例。您可以根据实例预热时间或其他应用程序需求配置时间长度。

使用简单扩展策略，但不使用其他扩展策略时，Amazon EC2 Auto Scaling 支持冷却时间。在 Auto Scaling 组使用简单扩展策略动态扩展后，它会等待冷却时间完成，然后再继续扩展活动。您可以使用与 Auto Scaling 组关联的默认冷却时间，也可以通过为策略指定冷却时间覆盖默认值。有关简单扩展的更多信息，请参阅[适用于 Amazon EC2 Auto Scaling 的简单扩展和分步扩展策略](#) (p. 81)。

在手动扩展 Auto Scaling 组时，默认设置是不等待冷却时间，但您可以覆盖默认设置并采用冷却时间。如果实例运行状况不佳，Auto Scaling 组不会等待冷却时间完成才替换运行状况不佳的实例。有关手动扩展的更多信息，请参阅[Amazon EC2 Auto Scaling 的手动扩展](#) (p. 65)。

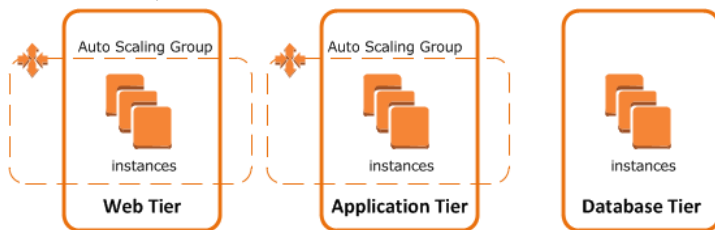
Amazon EC2 Auto Scaling 支持默认冷却时间和特定于扩展的冷却时间。

目录

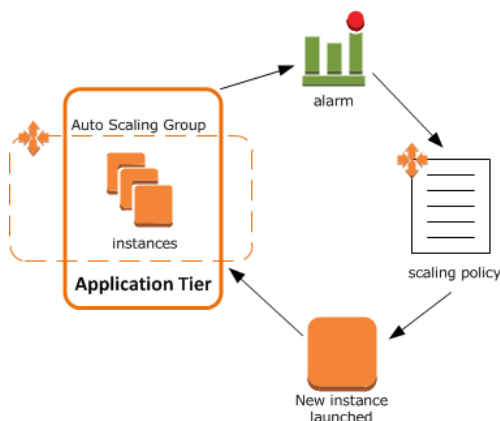
- [示例：冷却时间](#) (p. 93)
- [默认冷却时间](#) (p. 94)
- [特定于扩展的冷却时间](#) (p. 94)
- [冷却时间和多个实例](#) (p. 95)
- [冷却时间和生命周期挂钩](#) (p. 95)

示例：冷却时间

考虑以下情景：有一个在 AWS 中运行的 Web 应用程序。此 Web 应用程序包含三个基本层：Web、应用程序和数据库。为了确保应用程序始终拥有足够的资源来满足流量需求，您创建了两个 Auto Scaling 组：一个用于 Web 层，一个用于应用程序层。



为了帮助确保用于应用程序层的 Auto Scaling 组具有适当数量的 EC2 实例，请[创建一个 CloudWatch 警报](#) (p. 121)，每当实例的 CPUUtilization 指标超过 90% 时便会扩展。当警报发生时，Auto Scaling 组会启动并配置其他实例。



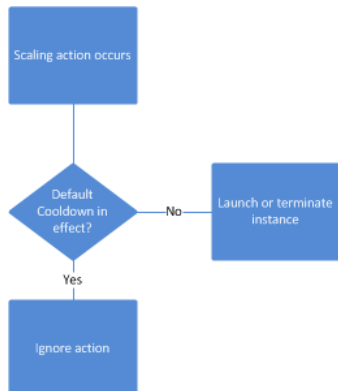
在实例投入使用之前，这些实例使用配置脚本安装和配置软件。因此，实例从启动到完全投入使用大约需要两到三分钟的时间。实际时间取决于诸多因素，如实例大小和是否有启动脚本要完成等。

现在出现了流量高峰，从而导致 CloudWatch 警报触发。该警报触发时，Auto Scaling 组会启动一个实例来帮助处理增加的需求。但是存在一个问题：该实例需要几分钟的时间才能启动。在此期间，CloudWatch 警报可能会继续触发，从而导致 Auto Scaling 组在警报每次出现时都另外启动一个实例。

但是，有了冷却时间，Auto Scaling 组在启动一个实例后，将暂停所有简单扩展策略或手动扩展引起的扩展活动，直至经过了指定时间量。（默认值为 300 秒。）这样，新启动的实例有时间开始处理应用程序流量。冷却时间过后，所有暂停的扩展操作都会恢复。如果 CloudWatch 警报再次触发，则 Auto Scaling 组将启动另一个实例，而冷却时间也会再次生效。不过，如果新增的实例足以将 CPU 使用率降为正常水平，则该组会保持其当前大小。

默认冷却时间

创建 Auto Scaling 组时会应用默认冷却时间。其默认值为 300 秒。此冷却时间自动应用于简单扩展策略的任何扩展活动，您可以选择请求将其应用于您的手动扩展活动。



在创建 Auto Scaling 组时，您可以使用 AWS 管理控制台、[create-auto-scaling-group](#) 命令 (AWS CLI) 或 [CreateAutoScalingGroup](#) API 操作配置默认冷却时间。

您可以随时使用 AWS 管理控制台、[update-auto-scaling-group](#) 命令 (AWS CLI) 或 [UpdateAutoScalingGroup](#) API 操作更改默认冷却时间。

特定于扩展的冷却时间

除了指定 Auto Scaling 组的默认冷却时间外，您还可以创建适用于特定简单扩展策略或手动扩展的冷却时间。特定于扩展的冷却时间会覆盖默认冷却时间。

特定于扩展的冷却时间常用于缩减策略，即根据特定条件或指标终止实例的策略。由于此策略是要终止实例，Amazon EC2 Auto Scaling 需要较短的时间来确定是否终止其他实例。300 秒的默认冷却时间太长，在这种情况下，为扩展策略使用特定于扩展的 180 秒冷却时间可以提高扩展组的速度并降低成本。

您可以使用 AWS 管理控制台、`put-scaling-policy` 命令 (AWS CLI) 或 `PutScalingPolicy` API 操作创建扩展特定的冷却时间。

冷却时间和多个实例

前面部分提供的示例演示了冷却时间在启动或终止单个实例时对 Auto Scaling 组的影响方式。不过，Auto Scaling 组经常一次启动多个实例。例如，您可能会选择让 Auto Scaling 组在达到特定指标阈值时启动三个实例。

对于多个实例，冷却时间（无论是默认冷却时间还是特定于扩展的冷却时间）会在最后一个实例启动时开始生效。

冷却时间和生命周期挂钩

您可以向 Auto Scaling 组添加生命周期挂钩。通过使用这些挂钩，您可以控制实例如何在 Auto Scaling 组中启动和终止。您可以在实例投入使用或终止之前对实例执行操作。

生命周期挂钩可以影响针对 Auto Scaling 组、手动扩展或简单扩展策略配置的任何冷却时间的行为。冷却时间在实例退出等待状态之后才会开始。

控制在缩小过程中终止哪些 Auto Scaling 实例

对于每个 Auto Scaling 组，您可以控制它何时向网络架构添加实例（称为扩展）或从中删除实例（称为缩减）。您可以通过附加和分离实例手动扩展组大小，或者可以使用扩展策略自动执行该过程。

要指定在缩减时首先终止哪一实例，请为 Auto Scaling 组配置终止策略。

在执行自动缩减时，您也可以使用实例保护来防止特定实例被终止。对于 Auto Scaling 组中的实例，使用 Amazon EC2 Auto Scaling 功能在发生缩减事件时保护实例。如果要防止您的实例被意外终止，请使用 Amazon EC2 终止保护。

Note

对于 [Auto Scaling 组具有多个实例类型和购买选项 \(p. 37\)](#)，缩减操作会考虑按需基数以及为按需和 Spot 实例设置的百分比。在组缩减时，Auto Scaling 组首先标识它将终止的两种实例类型。然后，它会选择有包含最多该类型实例并应用默认或自定义终止策略的可用区。

目录

- [默认终止策略 \(p. 95\)](#)
- [自定义终止策略 \(p. 97\)](#)
- [实例保护 \(p. 98\)](#)

默认终止策略

本节描述 Auto Scaling 在发生缩减事件时使用的默认终止策略。默认终止策略旨在帮助确保在可用区之间平均分配实例，以获得高可用性。默认策略会保留通用性和灵活性，旨在满足各种场景的需要。

使用默认终止策略时，Auto Scaling 组的行为如下所示：

1. 确定哪些可用区包含最多实例，并且至少有一个实例不受缩减保护。

如果可以从可用区中选择多个不受保护的实例，可以根据以下标准（按显示的顺序应用）选择要终止的实例。

2. [仅适用于 [Auto Scaling 组具有多个实例类型和购买选项 \(p. 37\)](#)]

确定要终止的实例，以便将剩余实例与要终止的按需或 Spot 实例的分配策略以及当前选择的实例类型保持一致。如果有一个此类实例，则终止该实例。否则，应用下一个条件。

3. [适用于使用启动模板的 Auto Scaling 组]

确定是否有任何实例使用最旧的启动模板。如果有一个此类实例，则终止该实例。（请注意存在一个例外情况：如果组最初使用启动配置。Amazon EC2 Auto Scaling 会首先终止使用启动配置的实例，然后再终止使用最旧启动模板的实例。）

4. [适用于使用启动配置的 Auto Scaling 组]

确定是否有任何实例使用最旧的启动配置。如果有一个此类实例，则终止该实例。

5. 在应用 2 到 4 中的所有条件后，如果要终止多个不受保护的实例，确定哪些实例最接近下一个计费小时。如果有一个此类实例，则终止该实例。（通过终止最接近下一个计费小时的实例，将帮助您最大程度地使用按小时计费的实例。）

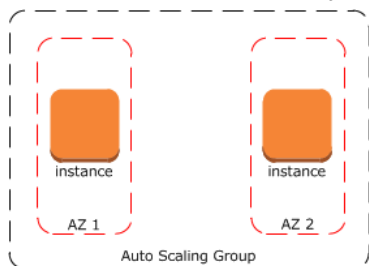
Note

使用按需实例时，您可以根据运行的实例以按小时或按分钟的方式为计算容量付费。如果您的 Auto Scaling 组使用 Amazon Linux 或 Ubuntu，则 EC2 的使用以秒为单位计费。有关更多信息，请参阅 [Amazon EC2 定价](#)。

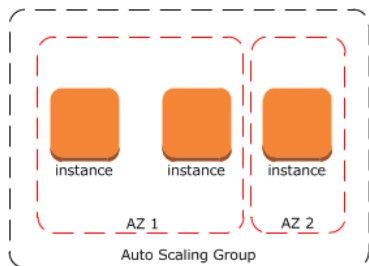
6. 如果多个未保护的实例最接近下一个计费小时，请随机选择其中的一个实例。

示例

设想一个使用启动配置的 Auto Scaling 组。它具有一个实例类型和两个可用区，所需的容量为两个实例，并且扩展策略在达到特定阈值时按一为增量增加和减少实例数。该组中的两个实例分布如下。



当达到扩展策略的阈值时，策略生效，Auto Scaling 组启动新实例。Auto Scaling 组中现在有三个实例，分布如下。



当达到缩减策略的阈值时，策略生效，Auto Scaling 组终止其中一个实例。如果您没有为该组分配特定的终止策略，则使用默认终止策略。它选择有两个实例的可用区，并终止从最旧启动配置启动的实例。如果这些实例是从同一启动配置启动的，则 Auto Scaling 组选择最接近下一个计费小时的实例并终止该实例。

有关 Amazon EC2 Auto Scaling 高可用性的更多信息，请参阅[在可用区之间分配实例 \(p. 5\)](#)。

自定义终止策略

为 Auto Scaling 组分配的默认终止策略通常足以适用于大多数情况。但是，您可以选择将默认策略替换为自定义策略。

当您自定义终止策略时，如果某个可用区的实例数多于该组使用的其他可用区，则您的终止策略将应用于不均衡可用区中的实例。如果组使用的可用区是均衡的，则终止策略将应用于该组的所有可用区。

Amazon EC2 Auto Scaling 支持以下自定义终止策略：

- **OldestInstance**。终止组中最旧的实例。当您将在 Auto Scaling 组中的实例升级为新的 EC2 实例类型，此选项十分有用。您可以逐渐将较旧类型的实例替换为较新类型的实例。
- **NewestInstance**。终止组中最新的实例。如果要测试新的启动配置但不想在生产中保留它时，此策略非常有用。
- **OldestLaunchConfiguration**。终止采用最旧启动配置的实例。如果要更新某个组并且逐步淘汰先前配置中的实例时，此策略非常有用。
- **ClosestToNextInstanceHour**。终止最接近下个计费小时的实例。此策略将帮助您最大程度地使用您的实例并管理 Amazon EC2 使用成本。
- **Default**。根据默认终止策略终止实例。如果该组有多个扩展策略，此策略非常有用。
- **OldestLaunchTemplate**。终止采用最旧启动模板的实例。利用此策略，会首先终止使用非当前启动模板的实例，然后终止使用当前启动模板的最旧版本的实例。如果要更新某个组并且逐步淘汰先前配置中的实例时，此策略非常有用。
- **AllocationStrategy**。终止 Auto Scaling 组中的实例，使剩余实例与所终止实例类型的分配策略匹配（Spot 实例或按需实例）。当您首选的实例类型发生变化时，可使用该策略。如果 Spot 分配策略是 `lowest-price`，则您可以逐渐在 N 个最低价 Spot 池中重新平衡分布 Spot 实例。如果 Spot 分配策略是 `capacity-optimized`，则您可以逐渐重新平衡具有更多可用 Spot 容量的 Spot 池之间 Spot 实例的分布。您也可以使用较高优先级类型的按需实例逐渐替代较低优先级类型的按需实例。

自定义终止策略（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Auto Scaling 组。
3. 选择 Auto Scaling 组。
4. 对于 Actions，选择 Edit。
5. 在 Details 选项卡上，找到 Termination Policies。选择一个或多个终止策略。如果您选择多个策略，请按照它们的应用顺序将其列出。如果您使用默认策略，请将其指定为列表中的最后一个策略。
6. 选择保存。

自定义终止策略 (AWS CLI)

使用以下命令之一：

- [create-auto-scaling-group](#)
- [update-auto-scaling-group](#)

您可以单独使用这些策略，或者将它们合并到策略列表中。例如，使用以下命令更新 Auto Scaling 组以首先使用 `OldestLaunchConfiguration` 策略，然后使用 `ClosestToNextInstanceHour` 策略：

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --termination-policies "OldestLaunchConfiguration" "ClosestToNextInstanceHour"
```


如果您使用 Default 终止策略，请将该策略设为终止策略列表中的最后一项。例如，`--termination-policies "OldestLaunchConfiguration" "Default"`。

实例保护

要控制 Auto Scaling 组在缩减时是否可以终止特定实例，请使用实例保护。您可以对 Auto Scaling 组或单个 Auto Scaling 实例启用实例保护设置。当 Auto Scaling 组启动实例时，实例会继承 Auto Scaling 组的实例保护设置。您可以随时更改 Auto Scaling 组或 Auto Scaling 实例的实例保护设置。

当实例状态为 `InService` 时，实例保护启动。当您分离受终止保护的实例时，其实例保护设置就会失效。再次将实例挂载到组时，它会继承组的当前实例保护设置。

如果缩减期间 Auto Scaling 组中的所有实例都受终止保护并且发生缩减事件，则其所需容量会减少。不过，Auto Scaling 组不能终止所需数量的实例，直到其实例保护设置被禁用。

实例保护并不能针对以下情况保护 Auto Scaling 实例：

- 通过 Amazon EC2 控制台、`terminate-instances` 命令或 `TerminateInstances` 操作手动终止。要保护 Auto Scaling 实例不被手动终止，请启用 Amazon EC2 终止保护。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[启用终止保护](#)。
- 实例未通过运行状况检查的情况下的运行状况检查更换。有关更多信息，请参阅[Auto Scaling 实例的运行状况检查](#) (p. 115)。要阻止 Amazon EC2 Auto Scaling 终止运行状况不佳的实例，请暂停 `ReplaceUnhealthy` 流程。有关更多信息，请参阅[暂停和恢复扩展流程](#) (p. 110)。
- Spot 实例中断。当 Spot 实例的容量不再可用，或 Spot 价格超过您的最高价，则将其终止。

任务

- [启用组的实例保护](#) (p. 98)
- [修改组的实例保护设置](#) (p. 99)
- [修改实例的实例保护设置](#) (p. 99)

启用组的实例保护

您可以在创建 Auto Scaling 组时启用实例保护。默认情况下，实例保护是禁用的。

启用实例保护（控制台）

在创建 Auto Scaling 组时，在配置 Auto Scaling 组详细信息页上，在高级详细信息下从实例保护中选择 `Protect From Scale In` 选项。

▼ Advanced Details

Load Balancing ⓘ	You currently don't have any load balancers Learn about Elastic Load Balancing
Health Check Grace Period ⓘ	<input type="text" value="300"/> seconds
Monitoring ⓘ	Amazon EC2 Detailed Monitoring metrics, which are provided at 1 minute frequency, are not enabled for the launch configuration my-ic. Instances launched from it will use Basic Monitoring metrics, provided at 5 minute frequency. Learn more
Instance Protection ⓘ	<div><input type="text" value="Protect From Scale In"/></div>

启用实例保护 (AWS CLI)

可以使用以下 `create-auto-scaling-group` 命令启用实例保护：

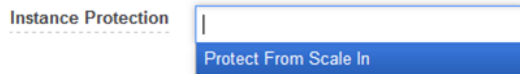
```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg --new-instances-protected-from-scale-in ...
```

修改组的实例保护设置

您可以启用或禁用 Auto Scaling 组的实例保护设置。

更改组的实例保护设置（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Auto Scaling 组。
3. 选择 Auto Scaling 组。
4. 在 Details 选项卡上，选择 Edit。
5. 对于实例保护，请选择保护免于缩减。



6. 选择保存。

更改组的实例保护设置 (AWS CLI)

可以使用以下 `update-auto-scaling-group` 命令为指定的 Auto Scaling 组启用实例保护：

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --new-instances-protected-from-scale-in
```

使用以下命令可对指定的组禁用实例保护：

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --no-new-instances-protected-from-scale-in
```

修改实例的实例保护设置

默认情况下，实例从其 Auto Scaling 组获取其实例保护设置。但是，您随时可以对实例启用或禁用实例保护。

更改实例的实例保护设置（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Auto Scaling 组。
3. 选择 Auto Scaling 组。
4. 在 Instances 选项卡上，选择实例。
5. 要启用实例保护，请依次选择 Actions、Instance Protection 和 Set Scale In Protection。系统提示时，选择 Set Scale In Protection。
6. 要禁用实例保护，请依次选择 Actions、Instance Protection 和 Remove Scale In Protection。系统提示时，选择删除缩减保护。

更改实例的实例保护设置 (AWS CLI)

使用以下 `set-instance-protection` 命令为指定的实例启用实例保护：

```
aws autoscaling set-instance-protection --instance-ids i-5f2e8a0d --auto-scaling-group-name my-asg --protected-from-scale-in
```

使用以下命令可对指定实例禁用实例保护：

```
aws autoscaling set-instance-protection --instance-ids i-5f2e8a0d --auto-scaling-group-name my-asg --no-protected-from-scale-in
```

Amazon EC2 Auto Scaling 生命周期挂钩

生命周期挂钩使您能够在 Auto Scaling 组启动或终止实例时通过暂停实例执行自定义操作。当实例被暂停后，它将保持等待状态，直到您使用 `complete-lifecycle-action` CLI 命令或 `CompleteLifecycleAction` API 操作完成生命周期操作，或者超时时段（默认为 1 小时）结束。

例如，您新启动的实例已完成其启动序列，一个生命周期挂钩暂停该实例。该实例处于等待状态时，您可以在其上安装或配置软件，确保您的实例在开始接收流量前已完全准备好。有关使用生命周期挂钩的另一个示例是，在发生缩减事件时，终止实例要首先从负载均衡器中取消注册（如果 Auto Scaling 组由 Elastic Load Balancing 使用）。然后，生命周期挂钩会在将实例终止之前将其暂停。当实例处于等待状态时，您可以执行连接到实例并下载日志或其他数据等操作，然后再将该实例完全终止。

每个 Auto Scaling 组均可具有多个生命周期挂钩。但是，每个 Auto Scaling 组的挂钩数存在限制。有关更多信息，请参阅 [Amazon EC2 Auto Scaling 限制 \(p. 8\)](#)。

目录

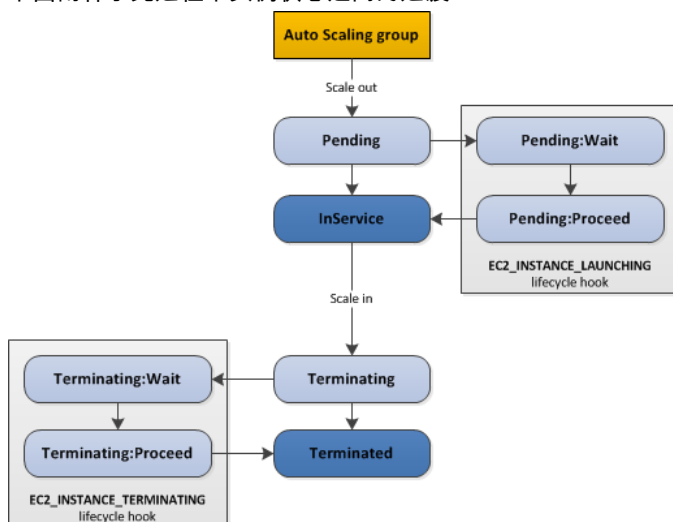
- [生命周期挂钩的工作方式 \(p. 100\)](#)
- [使用生命周期挂钩时的注意事项 \(p. 101\)](#)
- [准备通知 \(p. 102\)](#)
- [添加生命周期挂钩 \(p. 105\)](#)
- [完成生命周期挂钩 \(p. 105\)](#)
- [测试通知 \(p. 106\)](#)

生命周期挂钩的工作方式

在将生命周期挂钩添加到 Auto Scaling 组后，其工作方式如下：

1. Auto Scaling 组通过启动实例来响应扩展事件，通过终止实例来响应缩减事件。
2. 生命周期挂钩将实例置于等待状态（`Pending:Wait` 或 `Terminating:Wait`）。实例将暂停，直至您继续或者超时时段结束。
3. 您可以使用下面的一个或多个选项来执行自定义操作：
 - 定义一个 CloudWatch Events 目标以在生命周期操作发生时调用 Lambda 函数。在 Amazon EC2 Auto Scaling 将生命周期操作的事件提交到 CloudWatch Events 时，将调用 Lambda 函数。该事件包含有关实例是正在启动还是正在终止的信息，以及可用于控制生命周期操作的令牌。
 - 为生命周期挂钩定义一个通知目标。Amazon EC2 Auto Scaling 将向该通知目标发送消息。该消息包含有关实例是正在启动还是正在终止的信息，以及可用于控制生命周期操作的令牌。
 - 创建一个脚本，当实例启动时，该脚本将在实例上运行。该脚本可使用运行时所在的实例的 ID 来控制生命周期操作。
4. 默认情况下，实例将保持等待状态 1 小时，然后 Auto Scaling 组继续启动或终止过程（`Pending:Proceed` 或 `Terminating:Proceed`）。如果您需要更长时间，可通过记录检测信号来重新开始超时时段。如果您在超时时段结束前已完成此操作，则可以完成生命周期操作，从而继续启动或终止过程。

下图阐释了此过程中实例状态之间的过渡：



有关 Auto Scaling 组中实例的完整生命周期的更多信息，请参阅[Auto Scaling 生命周期](#) (p. 6)。

使用生命周期挂钩时的注意事项

向 Auto Scaling 组添加生命周期挂钩，使您可以对实例的启动和终止方式进行更多控制。下面是在向 Auto Scaling 组添加生命周期挂钩时为确保组按预期方式继续执行而要考虑的一些事项。

注意事项

- [使实例保持等待状态](#) (p. 101)
- [冷却时间和自定义操作](#) (p. 101)
- [运行状况检查宽限期](#) (p. 102)
- [生命周期操作结果](#) (p. 102)
- [Spot 实例](#) (p. 102)

使实例保持等待状态

实例可以在有限的时间里保持等待状态。默认值为 1 小时 (3600 秒)。您可以通过以下方式调整该时长：

- 在创建生命周期挂钩时，设置生命周期挂钩的检测信号超时。对于 [put-lifecycle-hook](#) 命令，请使用 `--heartbeat-timeout` 参数。对于 `PutLifecycleHook` 操作，请使用 `HeartbeatTimeout` 参数。
- 如果在超时时段结束前使用 [complete-lifecycle-action](#) 命令或 `CompleteLifecycleAction` 操作完成操作，则转到下一状态。
- 使用 [record-lifecycle-action-heartbeat](#) 命令或 `RecordLifecycleActionHeartbeat` 操作记录检测信号以推迟超时时段。这会将超时时间增加到您创建生命周期挂钩时指定的超时值。例如，如果超时值为 1 小时，而您在 30 分钟后调用该命令，则实例将继续保持等待状态 1 个小时 (总共为 90 分钟)。

您可以将实例保持在等待状态的最长时间是 48 小时或检测信号超时的 100 倍 (以较小者为准)。

冷却时间和自定义操作

当 Auto Scaling 组启动或终止简单扩展策略引起的实例时，[冷却时间](#) (p. 93)生效。冷却时间可帮助确保 Auto Scaling 组只根据需要启动或终止合适数量的实例。

考虑在实例启动时使用包含支持自定义操作的声明周期挂钩的 Auto Scaling 组。当应用程序需求增加时，该组会启动实例以增加容量。由于存在生命周期挂钩，实例会置于 `Pending:Wait` 状态，这意味着该实例尚无法用于处理流量。当实例进入等待状态时，简单扩展策略引起的扩展操作会暂停。当实例进入 `InService` 状态时，冷却时间开始。冷却时间过后，所有暂停的扩展操作都会恢复。

运行状况检查宽限期

如果添加一个生命周期挂钩，在生命周期挂钩操作完成并且实例进入 `InService` 状态后，运行状况检查宽限期才会开始。

生命周期操作结果

生命周期挂钩结束时，结果为 `ABANDON` 或 `CONTINUE`。

如果实例正在启动，则 `CONTINUE` 指示您的操作已成功，并且可将实例投入使用。否则，`ABANDON` 指示您的自定义操作未成功，并且可终止实例。

如果实例正在终止，则 `ABANDON` 和 `CONTINUE` 都允许终止实例。不过，`ABANDON` 将停止任何剩余操作（例如，其他生命周期挂钩），而 `CONTINUE` 将允许完成任何其他生命周期挂钩。

Spot 实例

您可以对 Spot 实例使用生命周期挂钩。但是，生命周期挂钩并不禁止在容量不再可用的情况下终止实例。此外，在 Spot 实例终止时，您仍需完成生命周期操作（使用 `complete-lifecycle-action` 命令或 `CompleteLifecycleAction` 操作）。

准备通知

您可以配置当实例进入等待状态时的通知，这能让您执行自定义操作。您可以使用 Amazon CloudWatch Events、Amazon SNS 或 Amazon SQS 接收通知。选择您喜好的选项。

Important

用于通知的 AWS 资源必须始终在您在其中创建生命周期挂钩的同一个 AWS 区域中创建。例如，如果您使用 Amazon SNS 通知，Amazon SNS 主题必须与您的生命周期挂钩位于同一个区域中。

或者，如果您拥有在实例启动时配置实例的脚本，则无需在生命周期操作发生时接收通知。如果您尚未这样做，请更新脚本，从实例元数据中检索实例的实例 ID。有关更多信息，请参阅[检索实例元数据](#)。

通知选项

- [使用 CloudWatch Events 将通知路由到 Lambda \(p. 102\)](#)
- [使用 Amazon SNS 接收通知 \(p. 103\)](#)
- [使用 Amazon SQS 接收通知 \(p. 104\)](#)

使用 CloudWatch Events 将通知路由到 Lambda

您可以使用 CloudWatch Events 设置一个目标以在生命周期操作发生时调用 Lambda 函数。

使用 CloudWatch Events 设置通知

1. 使用 [创建 Lambda 函数 \(p. 126\)](#) 中的步骤创建 Lambda 函数，并记下其 Amazon 资源名称 (ARN)。例如：`arn:aws:lambda:region:123456789012:function:my-function`。
2. 使用以下 `put-rule` 命令创建一个与生命周期操作匹配的 CloudWatch Events 规则。

```
aws events put-rule --name my-rule --event-pattern file://pattern.json --state ENABLED
```

实例启动生命周期操作的 pattern.json 为：

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-launch Lifecycle Action" ]
}
```

实例终止生命周期操作的 pattern.json 为：

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-terminate Lifecycle Action" ]
}
```

3. 使用以下 [add-permission](#) 命令为该规则授予调用 Lambda 函数的权限。此命令信任 CloudWatch Events 服务委托人 (events.amazonaws.com)，并将权限范围划定为指定规则。

```
aws lambda add-permission --function-name LogScheduledEvent --statement-id my-scheduled-event \
  --action 'lambda:InvokeFunction' --principal events.amazonaws.com --source-arn
arn:aws:events:region:123456789012:rule/my-scheduled-rule
```

4. 使用以下 [put-targets](#) 命令创建一个目标，它在执行生命周期操作时调用 Lambda 函数。

```
aws events put-targets --rule my-rule --targets
  Id=1,Arn=arn:aws:lambda:region:123456789012:function:my-function
```

5. 在遵循以上说明操作后，继续[添加生命周期挂钩](#) (p. 105)作为下一步。

当 Auto Scaling 组响应扩展或收缩事件时，它将实例置于等待状态。当实例处于等待状态时，Lambda 函数将被调用。有关事件数据的更多信息，请参阅 [Auto Scaling 事件](#) (p. 122)。

使用 Amazon SNS 接收通知

您可以使用 Amazon SNS 设置通知目标以便在生命周期操作开始时接收通知。

使用 Amazon SNS 设置通知

1. 使用以下 [create-topic](#) 命令创建 Amazon SNS 主题。有关更多信息，请参阅 Amazon Simple Notification Service 开发人员指南中的[创建主题](#)。

```
aws sns create-topic --name my-sns-topic
```

记下目标的 ARN (例如 `arn:aws:sns:region:123456789012:my-sns-topic`)。

2. 为 Amazon EC2 Auto Scaling 创建服务角色 (或代入角色)，您可以为其授予访问通知目标的权限。

要创建 IAM 角色并允许 Amazon EC2 Auto Scaling 担任该角色

- a. 通过以下网址打开 IAM 控制台：<https://console.aws.amazon.com/iam/>。
- b. 在导航窗格中，选择 Roles 和 Create new role。
- c. 在选择受信任实体的类型下，选择 AWS 服务。
- d. 在选择将使用此角色的服务下，选择 EC2 Auto Scaling。

- e. 在选择您的使用案例下，选择 EC2 Auto Scaling Notification Access (EC2 Auto Scaling 通知访问)，然后选择 Next:Permissions (下一步：权限)。
 - f. 选择 Next: Tags (下一步：标签)，(可选)通过附加密钥值对为角色添加元数据标签，然后选择 Next: Review (下一步：审核)。
 - g. 在审核页面上，键入角色的名称 (例如，my-notification-role)，然后选择创建角色。
 - h. 在角色页面中，选择刚刚创建的角色以打开摘要页面。记录角色 ARN。例如：arn:aws:iam::123456789012:role/my-notification-role。您将在下一步中创建生命周期挂钩时指定角色 ARN。
3. 在遵循以上说明操作后，继续[添加生命周期挂钩 \(p. 105\)](#)作为下一步。

当 Auto Scaling 组响应扩展或收缩事件时，它将实例置于等待状态。当实例处于等待状态时，将向通知目标发布消息。消息包含以下事件数据：

- LifecycleActionToken — 生命周期操作令牌。
- AccountId — AWS 账户 ID。
- AutoScalingGroupName — Auto Scaling 组的名称。
- LifecycleHookName — 生命周期挂钩的名称。
- EC2InstanceId — EC2 实例的 ID。
- LifecycleTransition — 生命周期挂钩类型。

例如：

```
Service: AWS Auto Scaling
Time: 2019-04-30T20:42:11.305Z
RequestId: 18b2ec17-3e9b-4c15-8024-ff2e8ce8786a
LifecycleActionToken: 71514b9d-6a40-4b26-8523-05e7ee35fa40
AccountId: 123456789012
AutoScalingGroupName: my-asg
LifecycleHookName: my-hook
EC2InstanceId: i-0598c7d356eba48d7
LifecycleTransition: autoscaling:EC2_INSTANCE_LAUNCHING
NotificationMetadata: null
```

使用 Amazon SQS 接收通知

您可以使用 Amazon SQS 设置通知目标以便在生命周期操作开始时接收通知。

Important

FIFO 队列与生命周期挂钩不兼容。

使用 Amazon SQS 设置通知

1. 使用 Amazon SQS 创建目标。有关更多信息，请参阅 Amazon Simple Queue Service 开发人员指南中的 [Amazon SQS 入门](#)。记下目标的 ARN (例如 arn:aws:sqs:region:123456789012:my-sqs-queue)。
2. 为 Amazon EC2 Auto Scaling 创建服务角色 (或代入角色)，您可以为其授予访问通知目标的权限。

要创建 IAM 角色并允许 Amazon EC2 Auto Scaling 担任该角色

- a. 通过以下网址打开 IAM 控制台：<https://console.aws.amazon.com/iam/>。
- b. 在导航窗格中，选择 Roles 和 Create new role。
- c. 在选择受信任实体的类型下，选择 AWS 服务。
- d. 在选择将使用此角色的服务下，选择 EC2 Auto Scaling。

- e. 在选择您的使用案例下，选择 EC2 Auto Scaling Notification Access (EC2 Auto Scaling 通知访问)，然后选择 Next:Permissions (下一步：权限)。
 - f. 选择 Next: Tags (下一步：标签)，(可选) 通过附加密钥值对为角色添加元数据标签，然后选择 Next: Review (下一步：审核)。
 - g. 在审核页面上，键入角色的名称 (例如，my-notification-role)，然后选择创建角色。
 - h. 在角色页面中，选择刚刚创建的角色以打开摘要页面。记录角色 ARN。例如：arn:aws:iam::**123456789012**:role/**my-notification-role**。您将在下一步中创建生命周期挂钩时指定角色 ARN。
3. 在遵循以上说明操作后，继续[添加生命周期挂钩](#) (p. 105)作为下一步。

当 Auto Scaling 组响应扩展或收缩事件时，它将实例置于等待状态。当实例处于等待状态时，将向通知目标发布消息。

添加生命周期挂钩

您可以使用 `put-lifecycle-hook` 命令创建生命周期挂钩。

要执行扩展操作，请使用以下命令。

```
aws autoscaling put-lifecycle-hook --lifecycle-hook-name my-hook --auto-scaling-group-name my-asg \  
  --lifecycle-transition autoscaling:EC2_INSTANCE_LAUNCHING
```

要执行缩减操作，请使用以下命令。

```
aws autoscaling put-lifecycle-hook --lifecycle-hook-name my-hook --auto-scaling-group-name my-asg \  
  --lifecycle-transition autoscaling:EC2_INSTANCE_TERMINATING
```

要使用 Amazon SNS 或 Amazon SQS 接收通知，您必须指定其他选项。如果您已按照前面的过程操作，已创建了通知目标和 IAM 角色。有关更多信息，请参阅 [准备通知](#) (p. 102)。

例如，添加以下选项可将 SNS 主题指定为通知目标。

```
--notification-target-arn arn:aws:sns:region:123456789012:my-sns-topic --role-arn  
arn:aws:iam::123456789012:role/my-notification-role
```

该主题将使用以下键/值对接收测试通知。

```
"Event": "autoscaling:TEST_NOTIFICATION"
```

完成生命周期挂钩

当 Auto Scaling 组响应扩展或缩减事件时，它会将实例置于等待状态并发送任意通知。在您完成生命周期挂钩后，它会继续启动或终止过程。

完成生命周期挂钩

1. 当实例处于等待状态时，您可以执行自定义操作。有关如何创建通知的信息，请参阅 [准备通知](#) (p. 102)。
2. 如果需要更多时间以完成自定义操作，请使用 `record-lifecycle-action-heartbeat` 命令重新开始超时时段，并将实例保持为等待状态。您可以指定在上一步中接收的生命周期操作令牌，如以下命令所示。

```
aws autoscaling record-lifecycle-action-heartbeat --lifecycle-hook-name my-launch-hook \
  --auto-scaling-group-name my-asg --lifecycle-action-
token bcd2f1b8-9a78-44d3-8a7a-4dd07d7cf635
```

或者，可以指定在上一步中接收的实例的 ID，如以下命令所示。

```
aws autoscaling record-lifecycle-action-heartbeat --lifecycle-hook-name my-launch-hook \
  --auto-scaling-group-name my-asg --instance-id i-1a2b3c4d
```

3. 如果在超时时段结束前完成自定义操作，请使用 `complete-lifecycle-action` 命令，以便 Auto Scaling 组可以继续启动或终止实例。您可以指定生命周期操作令牌，如以下命令所示。

```
aws autoscaling complete-lifecycle-action --lifecycle-action-result CONTINUE \
  --lifecycle-hook-name my-launch-hook --auto-scaling-group-name my-asg \
  --lifecycle-action-token bcd2f1b8-9a78-44d3-8a7a-4dd07d7cf635
```

或者，您可以指定实例的 ID，如以下命令所示。

```
aws autoscaling complete-lifecycle-action --lifecycle-action-result CONTINUE \
  --instance-id i-1a2b3c4d --lifecycle-hook-name my-launch-hook \
  --auto-scaling-group-name my-asg
```

测试通知

要为启动事件生成通知，请通过将 Auto Scaling 组的所需容量增加 1 来更新 Auto Scaling 组。您将在实例启动后的几分钟内收到通知。

更改所需容量（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Details 选项卡上，选择 Edit。
5. 对于 Desired，将当前值增加 1。如果此值超过 Max，还必须将 Max 的值增加 1。
6. 选择保存。
7. 在数分钟后，您将收到事件的通知。如果您不需要您为此测试启动的其他实例，则可以将 Desired 减少 1。在数分钟后，您将收到事件的通知。

临时从 Auto Scaling 组中删除实例

您可以将处于 InService 状态的实例置于 Standby 状态，更新实例或排查实例问题，然后将实例恢复运行状态。处于备用状态的实例仍是 Auto Scaling 组的一部分，但它们不会主动处理应用程序流量。

Important

您需为处于备用状态的实例付费。

例如，您可以随时更改 Auto Scaling 组的启动配置，Auto Scaling 组启动的任何后续实例将使用此配置。不过，Auto Scaling 组不会更新当前正在运行的实例。您可以终止这些实例，并让 Auto Scaling 组替换这些实例。或者，您可以将实例置于备用状态，更新软件，然后将实例恢复运行。

目录

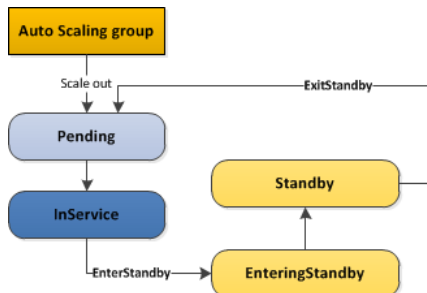
- [备用状态的工作方式 \(p. 107\)](#)
- [处于备用状态的实例的运行状况 \(p. 107\)](#)
- [使用临时删除实例 \(控制台\) \(p. 107\)](#)
- [临时删除实例 \(AWS CLI\) \(p. 108\)](#)

备用状态的工作方式

备用状态按如下方式工作以帮助您临时从 Auto Scaling 组中删除实例：

1. 将实例置于备用状态。实例保持此状态，直至您退出备用状态。
2. 如果已向 Auto Scaling 组附加负载均衡器或目标组，则将从该负载均衡器或目标组取消注册实例。
3. 在默认情况下，当您将一个实例置于备用状态时，指定作为所需容量的值会减少。这可防止在您将此实例置于备用状态时启动附加实例。或者，您可以指定不减少所需容量。如果您指定此选项，则 Auto Scaling 组会启动一个实例来替换处于备用状态的实例。这样做是为了帮助您在多个实例处于备用状态时保持应用程序的容量。
4. 您可以更新实例或排查实例的问题。
5. 可通过退出备用状态，将实例恢复运行状态。
6. 在您将处于备用状态的实例恢复运行后，所需容量会增加。如果在将实例置于备用状态时未减少容量，Auto Scaling 组将检测到您具有超过所需数量的实例。它应用终止策略以实际减小组的大小。有关更多信息，请参阅 [控制在缩小过程中终止哪些 Auto Scaling 实例 \(p. 95\)](#)。
7. 如果已向 Auto Scaling 组附加负载均衡器或目标组，则会将实例注册到该负载均衡器或目标组。

下图阐释了此过程中实例状态之间的过渡：



有关 Auto Scaling 组中实例的完整生命周期的更多信息，请参阅[Auto Scaling 生命周期 \(p. 6\)](#)。

处于备用状态的实例的运行状况

Amazon EC2 Auto Scaling 不对处于备用状态的实例执行运行状况检查。当实例处于备用状态时，其运行状况将反映您将其置于备用状态之前，实例具有的状态。Amazon EC2 Auto Scaling 不对实例执行运行状况检查，直至您将实例恢复运行。

例如，如果您将运行正常的实例置于备用状态，然后终止该实例，则 Amazon EC2 Auto Scaling 会继续将实例报告为运行正常。如果您将终止的实例恢复运行，则 Amazon EC2 Auto Scaling 将对实例执行运行状况检查，确定它运行状况不佳，并启动替换实例。

使用临时删除实例 (控制台)

以下步骤演示更新当前正在运行的实例的一般过程。

临时删除实例

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择 Auto Scaling 组。
4. 在 Instances 选项卡上，选择实例。
5. 选择 Actions 和 Set to Standby。
6. 在设置为备用页面上，选中该复选框可启动替代实例，取消选中可减少所需容量。选择设置为备用。
7. 您可以根据需要更新实例或排查实例的问题。当您完成后，请继续下一步以将实例恢复运行。
8. 选择实例，再选择 Actions 和 Set to InService。在设置为可用页面上，选择设置为可用。

临时删除实例 (AWS CLI)

以下步骤演示更新当前正在运行的实例的一般过程。

临时删除实例

1. 使用以下 `describe-auto-scaling-instances` 命令指定要更新的实例：

```
aws autoscaling describe-auto-scaling-instances
```

以下为响应示例：

```
{
  "AutoScalingInstances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "InService"
    },
    ...
  ]
}
```

2. 使用以下 `enter-standby` 命令将实例置于 Standby 状态。`--should-decrement-desired-capacity` 选项将减少所需容量以使 Auto Scaling 组不再启动替代实例。

```
aws autoscaling enter-standby --instance-ids i-05b4f7d5be44822a6 \
  --auto-scaling-group-name my-asg --should-decrement-desired-capacity
```

以下为响应示例：

```
{
  "Activities": [
    {
      "Description": "Moving EC2 instance to Standby: i-05b4f7d5be44822a6",
      "AutoScalingGroupName": "my-asg",

```

```
    "ActivityId": "3b1839fe-24b0-40d9-80ae-bcd883c2be32",
    "Details": "{\"Availability Zone\":\"us-west-2a\"}",
    "StartTime": "2014-12-15T21:31:26.150Z",
    "Progress": 50,
    "Cause": "At 2014-12-15T21:31:26Z instance i-05b4f7d5be44822a6 was moved to
standby
    in response to a user request, shrinking the capacity from 4 to 3.",
    "StatusCode": "InProgress"
  }
]
}
```

3. (可选) 使用以下 `describe-auto-scaling-instances` 命令验证实例是否处于 Standby。

```
aws autoscaling describe-auto-scaling-instances --instance-ids i-05b4f7d5be44822a6
```

以下为响应示例：注意，实例状态此时为 Standby。

```
{
  "AutoScalingInstances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "Standby"
    },
    ...
  ]
}
```

4. 您可以根据需要更新实例或排查实例的问题。当您完成后，请继续下一步以将实例恢复运行。
5. 使用以下 `exit-standby` 命令将实例恢复运行。

```
aws autoscaling exit-standby --instance-ids i-05b4f7d5be44822a6 --auto-scaling-group-name my-asg
```

以下为响应示例：

```
{
  "Activities": [
    {
      "Description": "Moving EC2 instance out of Standby: i-05b4f7d5be44822a6",
      "AutoScalingGroupName": "my-asg",
      "ActivityId": "db12b166-cdcc-4c54-8aac-08c5935f8389",
      "Details": "{\"Availability Zone\":\"us-west-2a\"}",
      "StartTime": "2014-12-15T21:46:14.678Z",
      "Progress": 30,
      "Cause": "At 2014-12-15T21:46:14Z instance i-05b4f7d5be44822a6 was moved
out of standby in
      response to a user request, increasing the capacity from 3 to 4.",
      "StatusCode": "PreInService"
    }
  ]
}
```

6. (可选) 使用以下 `describe-auto-scaling-instances` 命令验证实例是否已恢复运行。

```
aws autoscaling describe-auto-scaling-instances --instance-ids i-05b4f7d5be44822a6
```

以下为响应示例：请注意，实例状态为 `InService`。

```
{
  "AutoScalingInstances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "InService"
    },
    ...
  ]
}
```

暂停和恢复扩展流程

本主题介绍如何暂停然后恢复您的 Auto Scaling 组的一个或多个扩展进程。它还描述了选择使用 Amazon EC2 Auto Scaling 的暂停/恢复功能时要考虑的问题。

Important

如果您需要对实例进行问题排查或重启实例，请使用待机功能而不是暂停/恢复功能。有关更多信息，请参阅 [临时从 Auto Scaling 组中删除实例 \(p. 106\)](#)。在自动缩减时，您也可以使用实例保护功能来防止特定实例被终止。有关更多信息，请参阅 [实例保护 \(p. 98\)](#)。

除在启动时暂停外，Amazon EC2 Auto Scaling 也可以暂停多次启动实例失败的 Auto Scaling 组。这称为管理暂停。管理暂停最常用于符合以下条件的 Auto Scaling 组：连续尝试启动实例的时间超过 24 小时，但是未成功启动任何实例。您可以恢复由于管理原因被 Amazon EC2 Auto Scaling 暂停的进程。

目录

- [扩展流程 \(p. 110\)](#)
- [选择暂停 \(p. 111\)](#)
- [暂停和恢复扩展进程 \(控制台\) \(p. 113\)](#)
- [暂停和恢复扩展 \(AWS CLI\) \(p. 113\)](#)

扩展流程

对于 Amazon EC2 Auto Scaling，有两种主要进程类型：`Launch` 和 `Terminate`。`Launch` 进程将新的 Amazon EC2 实例添加到 Auto Scaling 组以增加其容量，而 `Terminate` 进程从组中删除 Amazon EC2 实例以减少其容量。

Amazon EC2 Auto Scaling 的另一个进程类型与特定扩展功能相关：

- `AddToLoadBalancer` — 在实例启动时，将其添加到附加的负载均衡器或目标组。

- **AlarmNotification** — 接受来自与组扩展策略关联的 CloudWatch 警报的通知。
- **AZRebalance** — 在组变得不均衡时，在指定的所有可用区中均匀平衡组中的 EC2 实例数，例如，之前不可用的可用区恢复到正常状态。有关更多信息，请参阅 [再平衡活动](#) (p. 6)。
- **HealthCheck** — 检查实例的运行状况，如果 Amazon EC2 或 Elastic Load Balancing 通知 Amazon EC2 Auto Scaling 实例运行状况不佳，则将对对应实例标记为运行状况不佳。此流程可覆盖您手动设置的实例运行状况状态。有关更多信息，请参阅 [Auto Scaling 实例的运行状况检查](#) (p. 115)。
- **ReplaceUnhealthy** — 终止被标记为运行状况不佳的实例，然后创建新实例以替换它们。
- **ScheduledActions** — 执行您创建的或 AWS Auto Scaling 的预测扩展功能创建的计划扩展操作。

选择暂停

可以单独暂停和恢复每个进程类型。本节提供在决定暂停扩展进程之前需要考虑的部分指导和行为。请记住，暂停单独进程可能会干扰其他进程。根据暂停进程的原因，您可能需要一起暂停多个进程。

下文说明暂停各个扩展类型时会发生什么。

Warning

如果您暂停 **Launch** 或 **Terminate** 进程类型，它可能导致其他进程类型无法正常工作。

Terminate

- 在暂停进程时，您的 Auto Scaling 组不会缩减到发出警报或执行计划操作。此外，以下进程会中断：
 - **AZRebalance** 仍处于活动状态，但不能正常运行。它可以启动新实例而不终止旧实例。这可能导致您的 Auto Scaling 组增加到比最大大小超出百分之十，因为在重新平衡活动期间允许短时间内发生这种情况。您的 Auto Scaling 组可以保持超出其最大大小，直到您恢复 **Terminate** 进程。当 **Terminate** 恢复后，如果 Auto Scaling 组在可用区之间不再平衡或者指定了不同的可用区，**AZRebalance** 逐渐重新平衡该组。
 - **ReplaceUnhealthy** 处于非活动状态但未 **HealthCheck**。当 **Terminate** 恢复后，**ReplaceUnhealthy** 进程将立即开始运行。如果任何实例在 **Terminate** 暂停期间被标记为运行状况不佳，将立即替换它们。

Launch

- 您的 Auto Scaling 组在暂停进程期间不会扩展到发出警报或执行计划操作。**AZRebalance** 会停止重新平衡组。**ReplaceUnhealthy** 会继续终止运行状况不佳的实例，但不会启动替换实例。在您恢复 **Launch** 后，通过以下方式完成重新平衡活动和运行状况检查替换：
 - 如果 Auto Scaling 组在可用区之间不再平衡，或者指定了不同的可用区，**AZRebalance** 会逐渐重新平衡组。
 - **ReplaceUnhealthy** 会立即替换它在 **Launch** 暂停期间终止的所有实例。

AddToLoadBalancer

- Amazon EC2 Auto Scaling 会启动实例，但不会将其添加到负载均衡器或目标组。在您恢复 **AddToLoadBalancer** 进程后，该进程也会在启动实例时将其添加到负载均衡器或目标组。不过，它不会添加在此流程暂停时启动的实例。您必须手动注册这些实例。

AlarmNotification

- 当超出 CloudWatch 警报阈值时，Amazon EC2 Auto Scaling 不执行扩展策略。暂停 **AlarmNotification** 可以暂时停止组扩展策略触发的扩展事件，而不删除扩展策略或其关联的 CloudWatch 警报。当您恢复 **AlarmNotification** 时，Amazon EC2 Auto Scaling 考虑具有当前违反的警报阈值的策略。

AZRebalance

- 您的 Auto Scaling 组不会在某些事件后尝试重新分配实例。如果发生扩展或缩减事件，扩展进程仍会尝试均衡可用区。例如，在扩展期间，它会在实例最少的可用区中启动实例。如果组变得不平衡，同时 AZRebalance 处于暂停状态并且您恢复它，Amazon EC2 Auto Scaling 会尝试重新平衡组。它先调用 Launch，然后调用 Terminate。

HealthCheck

- Amazon EC2 Auto Scaling 停止在 EC2 和 Elastic Load Balancing 运行状况检查后将实例标记为运行状况不佳。不过，您的自定义运行状况检查会继续正常运行。当您暂停 HealthCheck 后，在需要时可以手动设置组中实例的运行状况，并由 ReplaceUnhealthy 替换它们。

ReplaceUnhealthy

- Amazon EC2 Auto Scaling 停止替换标记为运行状况不佳的实例。未通过 EC2 或 Elastic Load Balancing 运行状况检查的实例仍会被标记为运行状况不佳。一旦您恢复 ReplaceUnhealthy 进程，Amazon EC2 Auto Scaling 会替换在此进程暂停时标记为运行状况不佳的实例。ReplaceUnhealthy 进程调用两种主要进程类型 — 先调用 Terminate，然后调用 Launch。

ScheduledActions

- 在暂停期间，Amazon EC2 Auto Scaling 不执行计划要运行的扩展操作。当您恢复 ScheduledActions 后，Amazon EC2 Auto Scaling 仅考虑尚未传递其执行时间的计划操作。

同时暂停启动和模板

当您同时暂停 Launch 和 Terminate 进程类型时，会发生以下情况：

- 您的 Auto Scaling 组无法启动扩展活动或保持其所需容量。
- 如果该组在可用区之间变得不平衡，Amazon EC2 Auto Scaling 不会尝试为您的 Auto Scaling 组在指定可用区之间重新均匀分布实例。
- 您的 Auto Scaling 组无法替换标记为运行状况不佳的实例。

当您恢复 Launch 和 Terminate 进程类型时，Amazon EC2 Auto Scaling 会替换在暂停进程期间标记为运行状况不佳的实例，并可能尝试重新平衡组。扩展活动也会恢复。

其他注意事项

在 Launch 和 Terminate 暂停时，某些外部操作可能会受影响。

- Spot 实例中断 — 如果 Terminate 已暂停而您的 Auto Scaling 组包含 Spot 实例，它们在 Spot 容量不再可用的情况下仍会终止。在 Launch 暂停期间，Amazon EC2 Auto Scaling 无法从另一个 Spot 实例池中启动替换实例，或者在其恢复可用后也无法从同一个 Spot 实例池中启动。
- 附加和分离实例 — 当 Launch 和 Terminate 暂停后，您可以将附加到您的 Auto Scaling 组的实例分离，但不能将新实例附加到组。要附加实例，必须首先恢复 Launch。

Note

如果分离实例后紧接着手动终止它，您可以改为调用 [terminate-instance-in-auto-scaling-group](#) CLI 命令。这将终止指定实例，并可以选择调整组的所需容量。此外，如果该 Auto Scaling 组用于生命周期挂钩，您为实例终止指定的自定义操作将在实例完全终止之前运行。

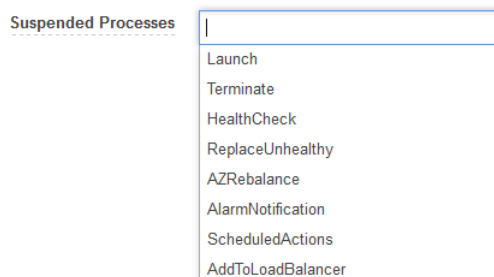
- 备用实例 — 在 Launch 暂停时，不能将 Standby 状态的实例恢复为运行。要将实例恢复运行，您必须首先恢复 Launch。

暂停和恢复扩展进程（控制台）

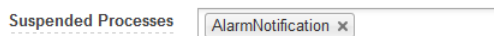
您可以暂停和恢复单个流程或所有流程。

暂停和恢复进程

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择 Auto Scaling 组。
4. 在 Details 选项卡上，选择 Edit。
5. 对于 Suspended Processes，选择要暂停的流程。



要恢复已暂停的流程，请从 Suspended Processes 中删除该流程。



6. 选择保存。

暂停和恢复扩展 (AWS CLI)

您可以暂停和恢复单个流程或所有流程。

暂停一个流程

将 `suspend-processes` 命令与 `--scaling-processes` 选项一起使用，如下所示：

```
aws autoscaling suspend-processes --auto-scaling-group-name my-asg --scaling-processes AlarmNotification
```

暂停所有流程

按以下方式使用 `suspend-processes` 命令（省略 `--scaling-processes` 选项）：

```
aws autoscaling suspend-processes --auto-scaling-group-name my-asg
```

恢复一个暂停的流程

按以下方式使用 `resume-processes` 命令：

```
aws autoscaling resume-processes --auto-scaling-group-name my-asg --scaling-processes AlarmNotification
```

恢复所有暂停的流程

按以下方式使用 `resume-processes` 命令 (省略 `--scaling-processes` 选项) :

```
aws autoscaling resume-processes --auto-scaling-group-name my-asg
```

监控您的 Auto Scaling 实例和组

您可以使用以下功能监控您的 Auto Scaling 实例和组。

运行状况检查

Amazon EC2 Auto Scaling 定期对 Auto Scaling 组中的实例执行运行状况检查并标识所有运行状况不佳的实例。可将 Auto Scaling 配置为使用 Amazon EC2 状态检查、Elastic Load Balancing 运行状况检查或自定义运行状况检查来确定实例的运行状况。有关更多信息，请参阅 [Auto Scaling 实例的运行状况检查 \(p. 115\)](#)。

CloudWatch 指标

Amazon EC2 Auto Scaling 将有关 Auto Scaling 组的数据点发布到 Amazon CloudWatch。通过使用 CloudWatch，您可以按一组有序的时间序列数据（称为指标）检索有关这些数据点的统计数据。您可使用这些指标来验证您的系统是否按预期运行。有关更多信息，请参阅 [使用 Amazon CloudWatch 监控 Auto Scaling 组和实例 \(p. 117\)](#)。

CloudWatch Events

当您的 Auto Scaling 组启动或终止实例时或者出现生命周期操作时，Amazon EC2 Auto Scaling 会将事件提交到 Amazon CloudWatch Events。这使您能够在事件发生时调用 Lambda 函数。有关更多信息，请参阅 [在 Auto Scaling 组扩展时获取 CloudWatch Events \(p. 122\)](#)。

SNS 通知

当您的 Auto Scaling 组启动或终止实例时，Amazon EC2 Auto Scaling 会发送 Amazon SNS 通知。有关更多信息，请参阅 [在 Auto Scaling 组扩展时获取 Amazon SNS 通知 \(p. 127\)](#)。

CloudTrail 日志

通过使用 AWS CloudTrail，可以跟踪您的 AWS 账户或代表您的 AWS 账户对 Amazon EC2 Auto Scaling API 进行的调用。CloudTrail 将信息存储在您指定的 Amazon S3 存储桶上的日志文件中。您可以使用这些日志文件监控 Auto Scaling 组的活动。日志包括发出的请求、请求来自的源 IP 地址、发出请求的用户、发出请求的时间，等等。有关更多信息，请参阅 [使用 AWS CloudTrail 记录 Amazon EC2 Auto Scaling API 调用 \(p. 132\)](#)。

Auto Scaling 实例的运行状况检查

Auto Scaling 实例的运行状况有正常和不佳两种。Auto Scaling 组中所有实例在开始启动时都处于良好运行状态。除非 Amazon EC2 Auto Scaling 收到实例运行状况不佳的通知，否则将认为它们运行状况良好。此通知可能来自以下一个或多个来源：Amazon EC2、Elastic Load Balancing 或自定义运行状况检查。

Amazon EC2 Auto Scaling 在将一个实例标记为运行状况不佳后，会计划替换该实例。如果您不想替换实例，可以暂停任何单个 Auto Scaling 组的运行状况检查流程。

实例运行状况

Amazon EC2 Auto Scaling 可以使用下列一项或多项检查来确定实例的运行状况：

- Amazon EC2 提供的用于识别可能影响实例的软硬件问题的状态检查。这包括实例状态检查和系统状态检查。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [状态检查类型](#)。
- 由 Elastic Load Balancing 提供的运行状况检查。
- 您的自定义运行状况检查。

EC2 状态检查是 Amazon EC2 Auto Scaling 的默认运行状况检查，并且不需要任何特殊配置。不过，您可以通过指定其他检查自定义您的 Auto Scaling 组执行的默认运行状况检查。有关更多信息，请参阅 [向 Auto Scaling 组添加 Elastic Load Balancing 运行状况检查 \(p. 56\)](#) 和 [自定义运行状况检查 \(p. 117\)](#)。

确定实例运行状况

一个实例在经过完全配置并通过初始运行状况检查后，将被 Amazon EC2 Auto Scaling 视为正常并进入 InService 状态。Amazon EC2 Auto Scaling 定期对 Auto Scaling 组中的实例执行运行状况检查并标识所有运行状况不佳的实例。

Amazon EC2 Auto Scaling 运行状况检查使用 Amazon EC2 状态检查的结果确定实例的运行状况。如果实例状态是除 running 以外的其他任何状态，或者系统状态为 impaired，则 Amazon EC2 Auto Scaling 认为实例运行状况不佳并启动替换实例。这包括当实例为任何以下状态时：

- stopping
- stopped
- terminating
- terminated

如果将一个负载均衡器或目标组附加到 Auto Scaling 组，您可以对该组进行配置，以便在 Elastic Load Balancing 报告实例为 OutOfService 时将其标记为运行状况不佳。如果为负载均衡器启用了连接耗尽，则在由于扩展事件或运行状况检查替换而终止实例之前，Amazon EC2 Auto Scaling 会等待进行中的请求完成或等待最大超时过期（以先到者为准）。如果将 Auto Scaling 组配置为使用 Elastic Load Balancing 运行状况检查时，Amazon EC2 Auto Scaling 通过检查 EC2 状态检查和 Elastic Load Balancing 运行状况检查来确定实例的运行状况。有关更多信息，请参阅 [向 Auto Scaling 组添加 Elastic Load Balancing 运行状况检查 \(p. 56\)](#)。

如果您有自定义运行状况检查，则可以将运行状况检查的信息发送给 Amazon EC2 Auto Scaling，以便 Amazon EC2 Auto Scaling 能够利用此信息。例如，如果您确定实例未按预期运行，则可将实例的运行状况设置为 Unhealthy。Amazon EC2 Auto Scaling 下次对实例执行运行状况检查时，它将确定实例运行状况不佳，然后启动替换实例。

运行状况检查宽限期

通常，刚刚投入使用的 Auto Scaling 实例需要进行预热，然后才能通过运行状况检查。Amazon EC2 Auto Scaling 等到运行状况检查宽限期结束后才检查实例的运行状况。Amazon EC2 状态检查和 Elastic Load Balancing 运行状况检查可以在运行状况检查宽限期过期之前完成。但在运行状况检查宽限期过期后，Amazon EC2 Auto Scaling 才会执行这些检查。为了给实例提供足够的预热时间，请确保运行状况检查宽限期包含应用程序的预期启动时间。如果添加一个生命周期挂钩，在生命周期挂钩操作完成并且实例进入 InService 状态后，宽限期才会开始。

替换运行状况不佳的实例

在 Amazon EC2 或 Elastic Load Balancing 运行状况检查中将实例标记为运行状况不佳之后，将立即安排对它进行替换。它不会自动恢复良好的运行状况。您可以手动进行干预，方法是调用 [SetInstanceHealth](#) 操作（或 set-instance-health 命令）以将实例的运行状况重新设置为良好。但是如果该实例已终止，将导致错误。因为将实例标记为运行状况不佳和实际终止该实例之间的间隔非常短，尝试使用 SetInstanceHealth 操作（或 set-instance-health 命令）将该实例重新设置为良好运行状况可能仅适用于已暂停的组。有关更多信息，请参阅 [暂停和恢复扩展流程 \(p. 110\)](#)。

Amazon EC2 Auto Scaling 将创建用于终止运行状况不佳的实例的新扩展活动，然后终止该实例。随后，另一个扩展活动将启动新实例以替代终止的实例。

实例终止后，任何关联的弹性 IP 地址都会取消关联，并且不会自动与新实例关联。必须手动将这些弹性 IP 地址与新实例关联。同样，实例终止后，其附加的 EBS 卷也会分离。必须手动将这些 EBS 卷附加到新实

例。有关更多信息，请参阅Amazon EC2 用户指南（适用于 Linux 实例）中的[取消关联弹性 IP 地址并将它与其他实例重新关联](#)和将 [Amazon EBS 卷附加到实例](#)。

自定义运行状况检查

如果您具有自己的运行状况检查系统，您可以将实例的运行状况信息从您的系统直接发送到 Amazon EC2 Auto Scaling。

可以使用以下 `set-instance-health` 命令将指定实例的运行状况设置为 Unhealthy。

```
aws autoscaling set-instance-health --instance-id i-123abc45d --health-status Unhealthy
```

使用以下 `describe-auto-scaling-groups` 命令验证实例状态是否为 Unhealthy。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg
```

下面是一个示例响应，说明实例的运行状况为 Unhealthy 并且实例正在终止。

```
{
  "AutoScalingGroups": [
    {
      ...
      "Instances": [
        {
          "InstanceId": "i-123abc45d",
          "AvailabilityZone": "us-west-2a",
          "HealthStatus": "Unhealthy",
          "LifecycleState": "Terminating",
          "LaunchConfigurationName": "my-lc"
        },
        ...
      ]
    }
  ]
}
```

使用 Amazon CloudWatch 监控 Auto Scaling 组和实例

利用 Amazon CloudWatch，您可以按一组有序的时间序列数据（称为指标）来检索统计数据。您可使用这些指标来验证您的系统是否按预期运行。

Amazon EC2 将向 CloudWatch 发送描述 Auto Scaling 实例的指标。这些指标对任何 EC2 实例（而不仅仅是 Auto Scaling 组中的实例）可用。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[实例指标](#)。

Auto Scaling 组可以向 CloudWatch 发送描述该组本身的指标。您必须启用这些指标。

目录

- [Auto Scaling 组指标 \(p. 118\)](#)
- [Auto Scaling 组指标的维度 \(p. 118\)](#)
- [启用 Auto Scaling 组指标 \(p. 118\)](#)
- [为 Auto Scaling 实例配置监控 \(p. 119\)](#)

- [查看 CloudWatch 指标 \(p. 120\)](#)
- [创建 Amazon CloudWatch 警报 \(p. 121\)](#)

Auto Scaling 组指标

AWS/AutoScaling 命名空间包括以下指标。

指标	说明
GroupMinSize	Auto Scaling 组的最小大小。
GroupMaxSize	Auto Scaling 组的最大大小。
GroupDesiredCapacity	Auto Scaling 组试图维护的实例数量。
GroupInServiceInstances	作为 Auto Scaling 组的一部分运行的实例数量。该指标不包括处于挂起或终止状态的实例。
GroupPendingInstances	处于挂起状态的实例数量。挂起的实例尚不可用。该指标不包括处于可用状态或终止状态的实例。
GroupStandbyInstances	处于 Standby 状态的实例数。处于此状态的实例仍在运行，但不能有效使用。
GroupTerminatingInstances	正处于终止过程中的实例的数量。该指标不包括处于可用状态或挂起状态的实例。
GroupTotalInstances	Auto Scaling 组中的实例总数。该指标用于标识处于可用状态、挂起状态和终止状态的实例的数量。

Auto Scaling 组指标的维度

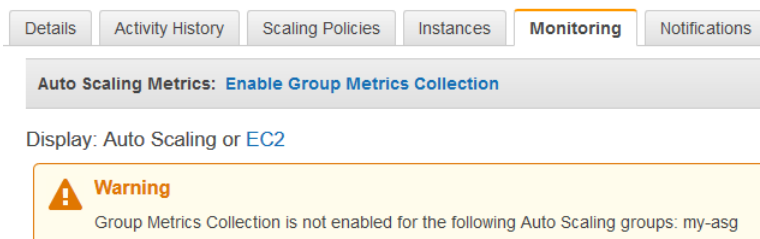
要按组名称筛选 Auto Scaling 组的指标，请使用 `AutoScalingGroupName` 维度。

启用 Auto Scaling 组指标

启用 Auto Scaling 组指标后，Auto Scaling 每分钟向 CloudWatch 发送采样数据。

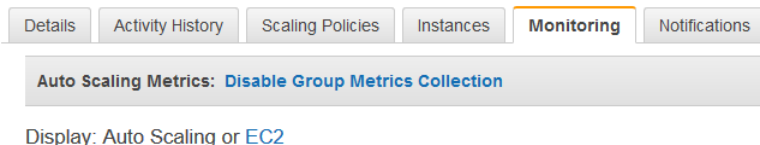
启用组指标（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Auto Scaling Groups。
3. 选择您的 Auto Scaling 组。
4. 在 Monitoring 选项卡上，为 Auto Scaling Metrics 选择 Enable Group Metrics Collection。如果您看不到此选项，请为 Display 选择 Auto Scaling。



禁用组指标 (控制台)

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Auto Scaling Groups。
3. 选择您的 Auto Scaling 组。
4. 在 Monitoring 选项卡上，为 Auto Scaling Metrics 选择 Disable Group Metrics Collection。如果您看不到此选项，请为 Display 选择 Auto Scaling。



启用组指标 (AWS CLI)

可以使用 `enable-metrics-collection` 命令启用一个或多个组指标。例如，以下命令可启用 GroupDesiredCapacity 指标。

```
aws autoscaling enable-metrics-collection --auto-scaling-group-name my-asg \
--metrics GroupDesiredCapacity --granularity "1Minute"
```

如果省略 `--metrics` 选项，则启用所有指标。

```
aws autoscaling enable-metrics-collection --auto-scaling-group-name my-asg \
--granularity "1Minute"
```

禁用组指标 (AWS CLI)

使用 `disable-metrics-collection` 命令。例如，以下命令可禁用所有 Auto Scaling 组指标：

```
aws autoscaling disable-metrics-collection --auto-scaling-group-name my-asg
```

为 Auto Scaling 实例配置监控

您可以使用启动配置或模板为 EC2 实例配置监控。每当启动实例时，都会启用监控：基本监控（5 分钟粒度）或详细监控（1 分钟粒度）。对于详细监控，将收取额外的费用。有关更多信息，请参阅 [Amazon CloudWatch 定价](#)。

Note

默认情况下，在创建启动模板或使用 AWS 管理控制台 创建启动配置时，将启用基本监控。在使用 AWS CLI 或开发工具包创建启动配置时，将默认启用详细监控。

要更改在新 EC2 实例上启用的监控类型，请更新启动模板或更新 Auto Scaling 组以使用新的启动配置。现有实例将继续使用以前启用的监控类型。要更新所有实例，请终止这些实例，以便 Auto Scaling 组替换这些实例，或使用 `monitor-instances` 和 `unmonitor-instances` 单独更新实例。

如果具有与 Auto Scaling 组关联的 CloudWatch 警报，请使用 `put-metric-alarm` 命令更新每个警报。使每个时段与监控类型匹配（基本监控为 300 秒，详细监控为 60 秒）。如果从详细监控更改为基本监控，但未更新警报以与 5 分钟时间段匹配，这些警报将继续每分钟检查一次统计数据。在每个 5 分钟时间段内，这些警报可能会在 4 分钟内找不到可用的数据。

配置 CloudWatch 监控 (控制台)

在使用 AWS 管理控制台创建启动配置时，在 Configure Details (配置详细信息) 页上，选择 Enable CloudWatch detailed monitoring (启用 CloudWatch 详细监控)。否则，将启用基本监控。有关更多信息，请参阅 [创建启动配置 \(p. 28\)](#)。

要使用 AWS 管理控制台为启动模板启用详细监控，请在高级详细信息部分中为监控选择启用。否则，将启用基本监控。有关更多信息，请参阅 [为 Auto Scaling 组创建启动模板 \(p. 22\)](#)。

配置 CloudWatch 监控 (AWS CLI)

对于启动配置，请将 [create-launch-configuration](#) 命令与 `--instance-monitoring` 选项一起使用。将此选项设置为 `true` 可启用详细监控，将此选项设置为 `false` 可启用基本监控。

```
--instance-monitoring Enabled=true
```

对于启动模板，请使用 [create-launch-template](#) 命令，并传递一个包含用于创建启动模板的信息的 JSON 文件。将监控属性设置为 `"Monitoring":{"Enabled":true}` 以启用详细监控，或设置为 `"Monitoring":{"Enabled":false}` 以启用基本监控。

查看 CloudWatch 指标

您可以使用 Amazon EC2 控制台查看 Auto Scaling 组和实例的 CloudWatch 指标。这些指标显示为监控图表。

或者，您可以使用 CloudWatch 控制台查看这些指标。

使用 Amazon EC2 控制台查看指标

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格中，选择 Auto Scaling Groups。
3. 选择您的 Auto Scaling 组。
4. 选择 Monitoring 选项卡。
5. (可选) 要按时间筛选结果，请从 Showing data for 中选择时间范围。
6. 要查看组的指标，请为 Display 选择 Auto Scaling。要获得单个指标的一个较大视图，请选择其图形。以下指标可用于组：
 - 最小组大小 — GroupMinSize
 - 最大组大小 — GroupMaxSize
 - 所需容量 — GroupDesiredCapacity
 - 在服务实例中 — GroupInServiceInstances
 - 待处理实例 — GroupPendingInstances
 - 备用实例 — GroupStandbyInstances
 - 正在终止实例 — GroupTerminatingInstances
 - 实例总数 — GroupTotalInstances
7. 要查看实例的指标，请为 Display 选择 EC2。要获得单个指标的一个较大视图，请选择其图形。以下指标可用于实例：
 - CPU 使用率 — CPUUtilization
 - 磁盘读取 — DiskReadBytes
 - 读磁盘操作 — DiskReadOps
 - 磁盘写入 — DiskWriteBytes
 - 写磁盘操作 — DiskWriteOps
 - 网络输入 — NetworkIn
 - 网络输出 — NetworkOut

- 状态检查失败 (任意) — `StatusCheckFailed`
- 状态检查失败 (实例) — `StatusCheckFailed_Instance`
- 状态检查失败 (系统) — `StatusCheckFailed_System`

使用 CloudWatch 控制台查看指标

有关更多信息，请参阅[按 Auto Scaling 组聚合统计数据](#)。

查看 CloudWatch 指标 (AWS CLI)

要查看所有 Auto Scaling 组的所有指标，请使用以下 `list-metrics` 命令：

```
aws cloudwatch list-metrics --namespace "AWS/AutoScaling"
```

要查看单个 Auto Scaling 组的指标，请指定 `AutoScalingGroupName` 维度，如下所示：

```
aws cloudwatch list-metrics --namespace "AWS/AutoScaling" --dimensions  
Name=AutoScalingGroupName,Value=my-asg
```

要查看适用于所有 Auto Scaling 组的单个指标，请指定该指标的名称，如下所示：

```
aws cloudwatch list-metrics --namespace "AWS/AutoScaling" --metric-name  
GroupDesiredCapacity
```

创建 Amazon CloudWatch 警报

监控指标的一个目的是验证您的应用程序是否正在按预期运行。如果指标超出了您认为可接受的阈值，则可以让 CloudWatch 警报触发某种操作。您可以指定 CloudWatch 支持的任何警报操作。

可以通过标识要监控的指标来配置警报。例如，您可以配置警报来监控 Auto Scaling 组中 EC2 实例的平均 CPU 使用率。该操作在满足以下条件时向您发送通知：在您指定的连续周期内，组的平均 CPU 使用率超过了指定的阈值。例如，如果在连续的 4 个 1 分钟周期内，指标量达到或高于 70%。

有关更多信息，请参阅 Amazon CloudWatch 用户指南 中的[使用 Amazon CloudWatch 警报](#)。

创建基于平均 CPU 使用率的 CloudWatch 警报

1. 通过以下网址打开 CloudWatch 控制台：<https://console.aws.amazon.com/cloudwatch/>。
2. 在导航窗格中，依次选择 Alarms 和 Create Alarm。
3. 选择选择指标，然后选择 EC2。
4. 选择用于筛选结果的指标类别。要查看 Auto Scaling 组指标，请选择 By Auto Scaling Group (按 Auto Scaling 组)。如果您只希望针对单个实例的指标，请选择每个实例的指标。
5. 按以下所示选择指标：
 - a. 选择包含要在其上创建警报的 Auto Scaling 组或实例的行，然后选择 CPUUtilization 指标。
 - b. 选择 Graphed metrics 选项卡。
 - c. 对于统计数据，选择平均值。
 - d. 在周期下，选择警报的评估周期，例如 1 分钟。评估警报时，每个时间段聚合到一个数据点。

Note

周期越短，创建的警报越敏感。

- e. 选择选择指标。

6. 在条件下，通过定义阈值条件来定义警报。例如，您可以定义一个阈值，在指标值大于或等于 70% 时触发警报。
7. 在其他配置下，对于触发警报的数据点数，指定必须有多少个数据点（评估周期）必须处于 ALARM 状态才会触发警报，例如，全部 4 个数据点。这将创建一个警报，如果多个连续周期超出阈值，该警报将进入 ALARM（警报）状态。
8. 对于缺失数据处理，请选择其中一个选项。对于持续报告数据的指标，例如 CPUUtilization，您可能希望选择 将缺失的数据作为不良（超出阈值）处理，因为缺失的数据点可能表明存在问题。有关更多信息，请参阅 Amazon CloudWatch 用户指南 中的 [配置 CloudWatch 警报处理缺少数据的方式](#)。
9. 选择 Next。
10. 在配置操作下，定义要执行的操作。
11. 选择 Next。
12. 在添加描述下，输入警报的名称和描述，然后选择下一步。
13. 选择 Create Alarm。

在 Auto Scaling 组扩展时获取 CloudWatch Events

您可以使用 Amazon CloudWatch Events 自动执行您的 AWS 服务并自动响应系统事件，例如应用程序可用性问题或资源更改。AWS 服务中的事件将实时传输到 CloudWatch Events。您可以编写简单规则来指示您关注的事件，并指示要在事件匹配规则时执行的自动化操作。

CloudWatch Events 允许您设置接收 JSON 格式的事件的各种目标，例如 Lambda 函数或 Amazon SNS 主题。有关更多信息，请参阅 [Amazon CloudWatch Events 用户指南](#)。

了解 Amazon EC2 Auto Scaling 何时启动或终止 Auto Scaling 组中的 EC2 实例是非常有用的。可以将 Amazon EC2 Auto Scaling 配置为在 Auto Scaling 组扩展时将事件发送到 CloudWatch Events。

Note

在 Spot 实例即将被 Amazon EC2 回收前，您也会收到两分钟警告。有关更多信息，请参阅适用于 Linux 实例的 Amazon EC2 用户指南 中的 [Spot 实例中断通知](#)。

目录

- [Auto Scaling 事件 \(p. 122\)](#)
- [创建 Lambda 函数 \(p. 126\)](#)
- [将事件路由到 Lambda 函数 \(p. 126\)](#)

Auto Scaling 事件

Amazon EC2 Auto Scaling 支持在发生以下事件时向 CloudWatch Events 发送通知：

- [EC2 实例启动生命周期操作 \(p. 122\)](#)
- [EC2 实例启动成功 \(p. 123\)](#)
- [EC2 实例启动失败 \(p. 124\)](#)
- [EC2 实例终止生命周期操作 \(p. 124\)](#)
- [EC2 实例终止成功 \(p. 125\)](#)
- [EC2 实例终止失败 \(p. 125\)](#)

EC2 实例启动生命周期操作

由于生命周期挂钩，Amazon EC2 Auto Scaling 已将实例移至 Pending:Wait 状态。

事件数据

以下是此事件的示例数据。

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance-launch Lifecycle Action",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "LifecycleActionToken": "87654321-4321-4321-4321-210987654321",
    "AutoScalingGroupName": "my-asg",
    "LifecycleHookName": "my-lifecycle-hook",
    "EC2InstanceId": "i-1234567890abcdef0",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_LAUNCHING",
    "NotificationMetadata": "additional-info"
  }
}
```

EC2 实例启动成功

Amazon EC2 Auto Scaling 已成功启动一个实例。

事件数据

以下是此事件的示例数据。

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance Launch Successful",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn",
    "instance-arn"
  ],
  "detail": {
    "StatusCode": "InProgress",
    "Description": "Launching a new EC2 instance: i-12345678",
    "AutoScalingGroupName": "my-auto-scaling-group",
    "ActivityId": "87654321-4321-4321-4321-210987654321",
    "Details": {
      "Availability Zone": "us-west-2b",
      "Subnet ID": "subnet-12345678"
    },
    "RequestId": "12345678-1234-1234-1234-123456789012",
    "StatusMessage": "",
    "EndTime": "yyyy-mm-ddThh:mm:ssZ",
    "EC2InstanceId": "i-1234567890abcdef0",
    "StartTime": "yyyy-mm-ddThh:mm:ssZ",
    "Cause": "description-text"
  }
}
```

EC2 实例启动失败

Amazon EC2 Auto Scaling 未能启动实例。

事件数据

以下是此事件的示例数据。

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance Launch Unsuccessful",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn",
    "instance-arn"
  ],
  "detail": {
    "StatusCode": "Failed",
    "AutoScalingGroupName": "my-auto-scaling-group",
    "ActivityId": "87654321-4321-4321-4321-210987654321",
    "Details": {
      "Availability Zone": "us-west-2b",
      "Subnet ID": "subnet-12345678"
    },
    "RequestId": "12345678-1234-1234-1234-123456789012",
    "StatusMessage": "message-text",
    "EndTime": "yyyy-mm-ddThh:mm:ssZ",
    "EC2InstanceId": "i-1234567890abcdef0",
    "StartTime": "yyyy-mm-ddThh:mm:ssZ",
    "Cause": "description-text"
  }
}
```

EC2 实例终止生命周期操作

由于生命周期挂钩，Amazon EC2 Auto Scaling 已将实例移至 Terminating:Wait 状态。

事件数据

以下是此事件的示例数据。

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance-terminate Lifecycle Action",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "LifecycleActionToken": "87654321-4321-4321-4321-210987654321",
    "AutoScalingGroupName": "my-asg",
    "LifecycleHookName": "my-lifecycle-hook",
    "EC2InstanceId": "i-1234567890abcdef0",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_TERMINATING",
  }
}
```

```
    "NotificationMetadata": "additional-info"
  }
}
```

EC2 实例终止成功

Amazon EC2 Auto Scaling 已成功终止实例。

事件数据

以下是此事件的示例数据。

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance Terminate Successful",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn",
    "instance-arn"
  ],
  "detail": {
    "StatusCode": "InProgress",
    "Description": "Terminating EC2 instance: i-12345678",
    "AutoScalingGroupName": "my-auto-scaling-group",
    "ActivityId": "87654321-4321-4321-4321-210987654321",
    "Details": {
      "Availability Zone": "us-west-2b",
      "Subnet ID": "subnet-12345678"
    },
    "RequestId": "12345678-1234-1234-1234-123456789012",
    "StatusMessage": "",
    "EndTime": "yyyy-mm-ddThh:mm:ssZ",
    "EC2InstanceId": "i-1234567890abcdef0",
    "StartTime": "yyyy-mm-ddThh:mm:ssZ",
    "Cause": "description-text"
  }
}
```

EC2 实例终止失败

Amazon EC2 Auto Scaling 未能终止实例。

事件数据

以下是此事件的示例数据。

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance Terminate Unsuccessful",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn",
    "instance-arn"
  ]
}
```



```
],
"detail": {
  "StatusCode": "Failed",
  "AutoScalingGroupName": "my-auto-scaling-group",
  "ActivityId": "87654321-4321-4321-4321-210987654321",
  "Details": {
    "Availability Zone": "us-west-2b",
    "Subnet ID": "subnet-12345678"
  },
  "RequestId": "12345678-1234-1234-1234-123456789012",
  "StatusMessage": "message-text",
  "EndTime": "yyyy-mm-ddThh:mm:ssZ",
  "EC2InstanceId": "i-1234567890abcdef0",
  "StartTime": "yyyy-mm-ddThh:mm:ssZ",
  "Cause": "description-text"
}
}
```

创建 Lambda 函数

使用以下过程可创建 Lambda 函数来处理 Auto Scaling 事件。

创建 Lambda 函数

1. 通过以下网址打开 AWS Lambda 控制台：<https://console.aws.amazon.com/lambda/>。
2. 如果您是首次使用 Lambda，则会看到一个欢迎页面；请选择 Get Started Now (立即开始使用)；否则，请选择 Create a Lambda function (创建 Lambda 函数)。
3. 在 Select blueprint 页面上，为 Filter 键入 hello-world，然后选择 hello-world 蓝图。
4. 在 Configure triggers 页面上，选择 Next。
5. 在 Configure function 页面上，执行以下操作：
 - a. 键入 Lambda 函数的名称和说明。
 - b. 编辑 Lambda 函数的代码。例如，以下代码仅记录事件：

```
console.log('Loading function');

exports.handler = function(event, context) {
  console.log("AutoScalingEvent()");
  console.log("Event data:\n" + JSON.stringify(event, null, 4));
  context.succeed("...");
};
```

- c. 对于 Role，选择 Choose an existing role (如果您现在具有要使用的角色)，然后从 Existing role 中选择您的角色。或者，要创建新角色，请为 Role 选择其他选项之一，然后按照说明执行操作。
 - d. (可选) 对于 Advanced settings，进行所需的任何更改。
 - e. 选择 Next。
6. 在 Review 页面上，选择 Create function。

将事件路由到 Lambda 函数

使用以下过程将 Auto Scaling 事件路由到 Lambda 函数。

将事件路由到 Lambda 函数

1. 通过以下网址打开 CloudWatch 控制台：<https://console.aws.amazon.com/cloudwatch/>。
2. 在导航窗格上，选择 Events。

3. 选择 Create rule。
4. 对于 Event selector，选择 Auto Scaling 作为事件源。默认情况下，规则将应用于您的所有 Auto Scaling 组的所有 Auto Scaling 事件。或者，您可以选择特定的事件或特定的 Auto Scaling 组。
5. 对于 Targets，选择 Add target。选择 Lambda function (Lambda 函数) 作为目标类型，然后选择 Lambda 函数。
6. 选择 Configure details。
7. 对于 Rule definition，键入规则的名称和描述。
8. 选择 Create rule。

要测试您的规则，请更改 Auto Scaling 组的大小。如果您已对 Lambda 函数使用示例代码，则它会将事件记录到 CloudWatch Logs。

测试您的规则

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上，选择 Auto Scaling Groups (Auto Scaling 组)，然后选择您的 Auto Scaling 组。
3. 在 Details 选项卡上，选择 Edit。
4. 更改 Desired 的值，然后选择 Save。
5. 通过以下网址打开 CloudWatch 控制台：<https://console.aws.amazon.com/cloudwatch/>。
6. 在导航窗格上，选择 Logs。
7. 选择 Lambda 函数的日志组（例如，`/aws/lambda/my-function`）。
8. 选择日志流以查看事件数据。此时将显示数据，类似以下内容：

```
Event Data
▼ 2016-02-22T17:48:20.778Z ealfjqinxq6pwo9d Loading function
▼ START RequestId: 7560439b-d98c-11e5-932d-f52757e7aee0 Version: $LATEST
▼ 2016-02-22T17:48:20.813Z 7560439b-d98c-11e5-932d-f52757e7aee0 AutoScalingEvent()
▼ 2016-02-22T17:48:20.814Z 7560439b-d98c-11e5-932d-f52757e7aee0 Event data:
{
  "version": "0",
  "id": "df9b0c8c-89c8-4748-92cb-ac68a9029ada",
  "detail-type": "EC2 Instance Launch Successful",
  "source": "aws.autoscaling",
```

在 Auto Scaling 组扩展时获取 Amazon SNS 通知

了解 Amazon EC2 Auto Scaling 何时启动或终止 Auto Scaling 组中的 EC2 实例是非常有用的。Amazon SNS 协调和管理将通知传输或发送到订阅客户端或终端节点的过程。您可以将 Amazon EC2 Auto Scaling 配置为在 Auto Scaling 组扩展时发送 SNS 通知。

Amazon SNS 可以将通知作为 HTTP 或 HTTPS POST、电子邮件（SMTP，采用纯文本或 JSON 格式）或发布到 Amazon SQS 队列的消息进行传输。有关更多信息，请参阅 Amazon Simple Notification Service 开发人员指南 中的 [什么是 Amazon SNS](#)。

例如，如果将 Auto Scaling 组配置为使用 `autoscaling: EC2_INSTANCE_TERMINATE` 通知类型，并且您的 Auto Scaling 组终止了某个实例，则它会发送电子邮件通知。该电子邮件包含已终止实例的详细信息，如实例 ID 以及终止该实例的原因。

Tip

如果您愿意，可以使用 Amazon CloudWatch Events 将目标配置为在 Auto Scaling 组扩展时或在执行生命周期操作时调用 Lambda 函数。有关更多信息，请参阅 [在 Auto Scaling 组扩展时获取 CloudWatch Events \(p. 122\)](#)。

目录

- [SNS 通知 \(p. 128\)](#)
- [配置 Amazon SNS \(p. 128\)](#)
- [配置 Auto Scaling 组以发送通知 \(p. 129\)](#)
- [测试通知配置 \(p. 130\)](#)
- [验证是否收到扩展事件的通知 \(p. 130\)](#)
- [删除通知配置 \(p. 131\)](#)

SNS 通知

Amazon EC2 Auto Scaling 支持在发生以下事件时发送 Amazon SNS 通知。

事件	说明
autoscaling:EC2_INSTANCE_LAUNCH	实例启动成功
autoscaling:EC2_INSTANCE_LAUNCH_ERROR	实例启动失败
autoscaling:EC2_INSTANCE_TERMINATE	实例终止成功
autoscaling:EC2_INSTANCE_TERMINATE_ERROR	实例终止失败

消息包含以下信息：

- Event — 事件。
- AccountId — AWS 账户 ID。
- AutoScalingGroupName — Auto Scaling 组的名称。
- AutoScalingGroupARN — Auto Scaling 组的 ARN。
- EC2InstanceId — EC2 实例的 ID。

例如：

```
Service: AWS Auto Scaling
Time: 2016-09-30T19:00:36.414Z
RequestId: 4e6156f4-a9e2-4bda-a7fd-33f2ae528958
Event: autoscaling:EC2_INSTANCE_LAUNCH
AccountId: 123456789012
AutoScalingGroupName: my-asg
AutoScalingGroupARN: arn:aws:autoscaling:region:123456789012:autoScalingGroup...
ActivityId: 4e6156f4-a9e2-4bda-a7fd-33f2ae528958
Description: Launching a new EC2 instance: i-0598c7d356eba48d7
Cause: At 2016-09-30T18:59:38Z a user request update of AutoScalingGroup constraints to ...
StartTime: 2016-09-30T19:00:04.445Z
EndTime: 2016-09-30T19:00:36.414Z
StatusCode: InProgress
StatusMessage:
Progress: 50
EC2InstanceId: i-0598c7d356eba48d7
Details: {"Subnet ID":"subnet-id","Availability Zone":"zone"}
```

配置 Amazon SNS

要使用 Amazon SNS 发送电子邮件通知，必须先创建一个主题，然后用您的电子邮件地址订阅该主题。

创建一个 Amazon SNS 主题

SNS 主题是一个逻辑接入点，即 Auto Scaling 组用来发送通知的通信通道。您可通过为主题指定名称来创建主题。

有关更多信息，请参阅 Amazon Simple Notification Service 开发人员指南 中的 [创建主题](#)。

订阅 Amazon SNS 主题

要接收您的 Auto Scaling 组发送到该主题的通知，必须让一个终端节点订阅该主题。在此过程中，为 Endpoint (终端节点) 指定要用来接收来自 Amazon EC2 Auto Scaling 的通知的电子邮件地址。

有关更多信息，请参阅 Amazon Simple Notification Service 开发人员指南 中的 [订阅主题](#)。

确认 Amazon SNS 订阅

Amazon SNS 向在上一步骤中指定的电子邮件地址发送确认电子邮件。

确保打开来自 AWS 通知的电子邮件并选择链接以确认订阅，然后再继续执行下一步。

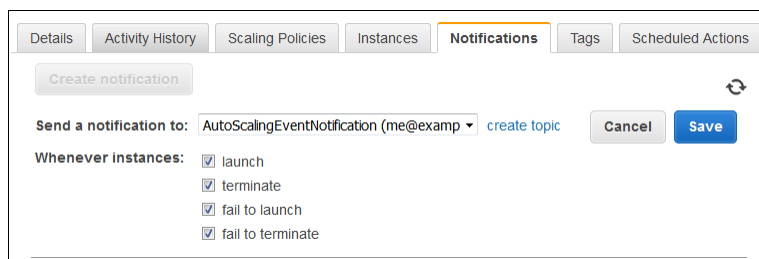
您将从 AWS 收到确认消息。Amazon SNS 现已配置为接收通知并将通知以电子邮件形式发送到您指定的电子邮件地址。

配置 Auto Scaling 组以发送通知

您可以配置 Auto Scaling 组，以便在发生扩展事件（例如，启动或终止实例）时向 Amazon SNS 发送通知。Amazon SNS 向您指定的电子邮件地址发送通知，通知中包含有关实例的信息。

为 Amazon SNS 组配置 Auto Scaling 通知（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Notifications 选项卡上，选择 Create notification。
5. 在 Create notifications 窗格上，执行以下操作：
 - a. 对于 Send a notification to:，选择您的 SNS 主题。
 - b. 对于 Whenever instances:，选择要为其发送通知的事件。
 - c. 选择 Save。



The screenshot shows the 'Create notification' dialog in the Amazon EC2 console. The 'Notifications' tab is selected. The 'Send a notification to:' field is set to 'AutoScalingEventNotification (me@examp)' with a 'create topic' link. The 'Whenever instances:' section has four checked options: 'launch', 'terminate', 'fail to launch', and 'fail to terminate'. There are 'Cancel' and 'Save' buttons at the bottom right.

为 Amazon SNS 组配置 Auto Scaling 通知 (AWS CLI)

使用以下 [put-notification-configuration](#) 命令。

```
aws autoscaling put-notification-configuration --auto-scaling-group-name my-asg --topic-arn arn --notification-types "autoscaling:EC2_INSTANCE_LAUNCH" "autoscaling:EC2_INSTANCE_TERMINATE"
```

测试通知配置

要为启动事件生成通知，请将 Auto Scaling 组的所需容量增加 1 以更新 Auto Scaling 组。Amazon EC2 Auto Scaling 将启动 EC2 实例，并且您在几分钟内收到电子邮件通知。

更改所需容量（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Details 选项卡上，选择 Edit。
5. 对于 Desired，将当前值增加 1。如果此值超过 Max，还必须将 Max 的值增加 1。
6. 选择 Save。
7. 在数分钟后，您将收到启动事件的通知电子邮件。如果您不需要您为此测试启动的其他实例，则可以将 Desired 减少 1。在数分钟后，您将收到终止事件的通知电子邮件。

更改所需容量 (AWS CLI)

使用以下 [set-desired-capacity](#) 命令。

```
aws autoscaling set-desired-capacity --auto-scaling-group-name my-asg --desired-capacity 2
```

验证是否收到扩展事件的通知

检查来自 Amazon SNS 的电子邮件消息并打开电子邮件。在收到 Auto Scaling 组的扩展事件通知后，您可以通过查看 Auto Scaling 组的描述来确认该扩展事件。您需要使用通知电子邮件中的信息，例如，启动或终止的实例的 ID。

验证 Auto Scaling 组已启动新实例（控制台）

1. 选择您的 Auto Scaling 组。
2. 在 Activity History 选项卡上，Status 列显示实例的当前状态。例如，如果通知指示实例已启动，则使用刷新按钮来验证启动活动的状态是否为 Successful。
3. 在 Instances 选项卡上，可以查看通知电子邮件中收到的 ID 的实例的当前 Lifecycle 状态。新实例启动后，其生命周期状态更改为 InService。

验证 Auto Scaling 组已启动新实例 (AWS CLI)

可以使用以下 [describe-auto-scaling-groups](#) 命令确认 Auto Scaling 组大小已更改。

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

下面的示例输出表明该组有两个实例。检查通知电子邮件中收到的 ID 的实例。

```
{
  "AutoScalingGroups": [
```

```
{
  "AutoScalingGroupARN": "arn",
  "HealthCheckGracePeriod": 0,
  "SuspendedProcesses": [],
  "DesiredCapacity": 2,
  "Tags": [],
  "EnabledMetrics": [],
  "LoadBalancerNames": [],
  "AutoScalingGroupName": "my-asg",
  "DefaultCooldown": 300,
  "MinSize": 1,
  "Instances": [
    {
      "InstanceId": "i-d95eb0d4",
      "AvailabilityZone": "us-west-2b",
      "HealthStatus": "Healthy",
      "LifecycleState": "InService",
      "LaunchConfigurationName": "my-lc"
    },
    {
      "InstanceId": "i-13d7dc1f",
      "AvailabilityZone": "us-west-2a",
      "HealthStatus": "Healthy",
      "LifecycleState": "InService",
      "LaunchConfigurationName": "my-lc"
    }
  ],
  "MaxSize": 5,
  "VPCZoneIdentifier": null,
  "TerminationPolicies": [
    "Default"
  ],
  "LaunchConfigurationName": "my-lc",
  "CreatedTime": "2015-03-01T16:12:35.608Z",
  "AvailabilityZones": [
    "us-west-2b",
    "us-west-2a"
  ],
  "HealthCheckType": "EC2"
}
```

删除通知配置

您可以随时删除 Amazon EC2 Auto Scaling 通知配置。

删除 Amazon EC2 Auto Scaling 通知配置（控制台）

1. 打开 Amazon EC2 控制台 <https://console.aws.amazon.com/ec2/>。
2. 在导航窗格上的 Auto Scaling 下，选择 Auto Scaling 组。
3. 选择您的 Auto Scaling 组。
4. 在 Notifications 选项卡上，选择通知旁边的 Delete。

删除 Amazon EC2 Auto Scaling 通知配置 (AWS CLI)

使用以下 delete-notification-configuration 命令。

```
aws autoscaling delete-notification-configuration --auto-scaling-group-name my-asg --topic-arn arn
```

有关删除 Amazon SNS 主题以及与 Auto Scaling 组关联的所有订阅的信息，请参阅 Amazon Simple Notification Service 开发人员指南 中的 [清除](#)。

使用 AWS CloudTrail 记录 Amazon EC2 Auto Scaling API 调用

Amazon EC2 Auto Scaling 与 AWS CloudTrail 集成在一起，后者是一项服务，它提供用户、角色或使用 Amazon EC2 Auto Scaling 的 AWS 服务执行的操作的记录。CloudTrail 将 Amazon EC2 Auto Scaling 的所有 API 调用作为事件进行捕获。捕获的调用包含来自 Amazon EC2 Auto Scaling 控制台的调用和对 Amazon EC2 Auto Scaling API 的代码调用。

如果您创建跟踪，则可以使 CloudTrail 事件持续传送到 Amazon S3 存储桶（包括 Amazon EC2 Auto Scaling 的事件）。如果您不配置跟踪，则仍可在 CloudTrail 控制台的 Event history（事件历史记录）中查看最新事件。通过使用 CloudTrail 收集的信息，您可以确定向 Amazon EC2 Auto Scaling 发出的请求、从中发出请求的 IP 地址、发出请求的用户、发出请求的时间以及其他详细信息。

要了解有关 CloudTrail 的更多信息，请参阅 [AWS CloudTrail User Guide](#)。

CloudTrail 中的 Amazon EC2 Auto Scaling 信息

在您创建 CloudTrail 账户时，即针对该账户启用了 AWS。Amazon EC2 Auto Scaling 中发生活动时，该活动将记录在 CloudTrail 事件中，并与其他 AWS 服务事件一同保存在 Event history（事件历史记录）中。您可以在 AWS 账户中查看、搜索和下载最新事件。有关更多信息，请参阅 [使用 CloudTrail 事件历史记录查看事件](#)。

要持续记录 AWS 账户中的事件（包括 Amazon EC2 Auto Scaling 的事件），请创建跟踪。通过跟踪，CloudTrail 可将日志文件传送到 Amazon S3 存储桶。默认情况下，在控制台中创建跟踪时，此跟踪应用于所有 AWS 区域。此跟踪在 AWS 分区中记录来自所有区域的事件，并将日志文件传送到您指定的 Amazon S3 存储桶。此外，您可以配置其他 AWS 服务，进一步分析在 CloudTrail 日志中收集的事件数据并采取行动。有关更多信息，请参阅下列内容：

- [创建跟踪概述](#)
- [CloudTrail 支持的服务和集成](#)
- [为 CloudTrail 配置 Amazon SNS 通知](#)
- [接收来自多个区域的 CloudTrail 日志文件和从多个账户接收 CloudTrail 日志文件](#)

所有 Amazon EC2 Auto Scaling 操作均由 CloudTrail 记录下来并记载到 [Amazon EC2 Auto Scaling API 参考](#) 中。例如，对 CreateLaunchConfiguration、DescribeAutoScalingGroup 和 UpdateAutoScalingGroup 操作进行的调用会在 CloudTrail 日志文件中生成条目。

每个事件或日志条目都包含有关生成请求的人员的信息。身份信息帮助您确定以下内容：

- 请求是使用根用户凭证还是 AWS Identity and Access Management (IAM) 用户凭证发出的。
- 请求是使用角色还是联合身份用户的临时安全凭证发出的。
- 请求是否由其他 AWS 服务发出。

有关更多信息，请参阅 [CloudTrail userIdentity 元素](#)。

了解 Amazon EC2 Auto Scaling 日志文件条目

跟踪是一种配置，可用于将事件作为日志文件传送到您指定的 Amazon S3 存储桶。CloudTrail 日志文件包含一个或多个日志条目。一个事件表示来自任何源的一个请求，包括有关所请求的操作、操作的日期和时间、

请求参数等方面的信息。CloudTrail 日志文件不是公用 API 调用的有序堆栈跟踪，因此它们不会以任何特定顺序显示。

下面的示例显示了一个 CloudTrail 日志条目，该条目说明了 CreateLaunchConfiguration 操作。

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "Root",
    "principalId": "123456789012",
    "arn": "arn:aws:iam::123456789012:root",
    "accountId": "123456789012",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
    "sessionContext": {
      "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2018-08-21T17:05:42Z"
      }
    }
  },
  "eventTime": "2018-08-21T17:07:49Z",
  "eventSource": "autoscaling.amazonaws.com",
  "eventName": "CreateLaunchConfiguration",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "192.0.2.0",
  "userAgent": "Coral/Jakarta",
  "requestParameters": {
    "ebsOptimized": false,
    "instanceMonitoring": {
      "enabled": false
    },
    "instanceType": "t2.micro",
    "keyName": "EC2-key-pair-oregon",
    "blockDeviceMappings": [
      {
        "deviceName": "/dev/xvda",
        "ebs": {
          "deleteOnTermination": true,
          "volumeSize": 8,
          "snapshotId": "snap-01676e0a2c3c7de9e",
          "volumeType": "gp2"
        }
      }
    ],
    "launchConfigurationName": "launch_configuration_1",
    "imageId": "ami-6cd6f714d79675a5",
    "securityGroups": [
      "sg-00c429965fd921483"
    ]
  },
  "responseElements": null,
  "requestID": "0737e2ea-fb2d-11e3-bfd8-99133058e7bb",
  "eventID": "3fcfb182-98f8-4744-bd45-b38835ab61cb",
  "eventType": "AwsApiCall",
  "recipientAccountId": "123456789012"
}
```

控制对您 Amazon EC2 Auto Scaling 资源的访问权

要访问 Amazon EC2 Auto Scaling，需要具有 AWS 可用于验证您的请求的凭证。这些凭证必须具有执行 Amazon EC2 Auto Scaling 操作的[权限](#)。

本主题提供了详细信息来说明 IAM 管理员如何使用 AWS Identity and Access Management (IAM) 控制哪些用户可执行 Amazon EC2 Auto Scaling 操作，从而对您的资源进行保护。

默认情况下，全新的 IAM 用户没有执行任何操作的权限。要授予调用 Amazon EC2 Auto Scaling 操作的权限，您可以将一个 IAM 策略附加到需要获得权限的 IAM 用户或组。

有关为用户授予 Amazon EC2 Auto Scaling 的完全访问权限的策略，请参阅[预定义的 AWS 托管策略 \(p. 137\)](#)。您也可以创建自己的策略来指定允许或拒绝的操作及资源，以及在允许或拒绝操作时所适用的条件。有关示例，请参阅[客户托管策略示例 \(p. 137\)](#)。但是，我们建议您首先阅读以下介绍性主题，这些主题讲解了管理 Amazon EC2 Auto Scaling 资源访问的基本概念和选项。

在策略中指定操作

在 IAM 策略中，您可以指定任何及所有 Amazon EC2 Auto Scaling 操作。有关更多信息，请参阅 Amazon EC2 Auto Scaling API 参考 中的[操作](#)。

要指定单个策略，您可以在操作名称中使用以下前缀：autoscaling:。

```
"Action": "autoscaling:CreateAutoScalingGroup"
```

要在单个策略中指定多项操作，请使用方括号将操作括起来并以逗号分隔。

```
"Action": [
    "autoscaling:CreateAutoScalingGroup",
    "autoscaling:UpdateAutoScalingGroup"
]
```

支持通配符。例如，您可以使用 autoscaling:* 指定所有 Amazon EC2 Auto Scaling 操作。

```
"Action": "autoscaling:*"
```

您还可以使用 Describe* 指定名称以 Describe 开头的所有操作。

```
"Action": "autoscaling:Describe*"
```

其他 IAM 权限

用户必须在 Amazon EC2 和 IAM 中具有额外的权限才能执行某些操作。您可以在 IAM 策略声明的 Action 元素中指定以下操作。

使用启动配置创建 Auto Scaling 组

- autoscaling:CreateAutoScalingGroup

- iam:CreateServiceLinkedRole

使用启动模板创建 Auto Scaling 组

- autoscaling:CreateAutoScalingGroup
- iam:CreateServiceLinkedRole
- ec2:RunInstances

更新使用启动模板的 Auto Scaling 组

- autoscaling:UpdateAutoScalingGroup
- ec2:RunInstances

创建启动配置

- autoscaling:CreateLaunchConfiguration
- ec2:DescribeImages
- ec2:DescribeInstances
- ec2:DescribeInstanceAttribute
- ec2:DescribeKeyPairs
- ec2:DescribeSecurityGroups
- ec2:DescribeSpotInstanceRequests
- ec2:DescribeVpcClassicLink

用户可能需要额外的权限才能创建或使用 Amazon EC2 资源，例如，从控制台中使用卷和管理安全组。有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[用于 Amazon EC2 控制台的策略示例](#)。有关创建和更新启动模板的权限的信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的[控制如何使用启动模板](#)。

此外，还可能需执行针对 CloudWatch、Elastic Load Balancing、IAM 和 Amazon SNS 的额外 API 操作。例如，需要执行 iam:PassRole 操作以使用[实例配置文件](#) (p. 147)。

指定资源

可以使用 IAM 策略控制对资源的访问。对于支持资源级权限的操作，您可以使用 Amazon 资源名称 (ARN) 标识策略适用的资源。

要指定 Auto Scaling 组，您必须指定其 ARN，如下所示：

```
"Resource":  
  "arn:aws:autoscaling:region:123456789012:autoScalingGroup:uuid:autoScalingGroupName/asg-name"
```

要指定启动配置，您必须指定其 ARN，如下所示：

```
"Resource":  
  "arn:aws:autoscaling:region:123456789012:launchConfiguration:uuid:launchConfigurationName/lc-name"
```

要使用 `CreateAutoScalingGroup` 操作指定 Auto Scaling 组，您必须将 UUID 替换为 *，如下所示：

```
"Resource":  
  "arn:aws:autoscaling:region:123456789012:autoScalingGroup:*:autoScalingGroupName/asg-name"
```

要使用 `CreateLaunchConfiguration` 操作指定启动配置，您必须将 UUID 替换为 *，如下所示：

```
"Resource":  
  "arn:aws:autoscaling:region:123456789012:launchConfiguration:*:launchConfigurationName/lc-name"
```

或者，如果不希望将特定资源作为目标，您也可以将 * 通配符作为资源。

```
"Resource": "*" 
```

可以在 IAM 策略中使用所有 Amazon EC2 Auto Scaling 操作以授予或拒绝用户使用该操作的权限。但是，并非所有 Amazon EC2 Auto Scaling 操作都支持资源级权限，这些权限允许指定可对其执行操作的资源。

对于不支持资源级权限的操作，您必须将 "*" 作为资源。

以下 Amazon EC2 Auto Scaling 操作不支持资源级权限：

- `DescribeAccountLimits`
- `DescribeAdjustmentTypes`
- `DescribeAutoScalingGroups`
- `DescribeAutoScalingInstances`
- `DescribeAutoScalingNotificationTypes`
- `DescribeLaunchConfigurations`
- `DescribeLifecycleHooks`
- `DescribeLifecycleHookTypes`
- `DescribeLoadBalancers`
- `DescribeLoadBalancerTargetGroups`
- `DescribeMetricCollectionTypes`
- `DescribeNotificationConfigurations`
- `DescribePolicies`
- `DescribeScalingActivities`
- `DescribeScalingProcessTypes`
- `DescribeScheduledActions`
- `DescribeTags`
- `DescribeTerminationPolicyTypes`

在策略中指定条件

对于支持资源级权限的操作，您可以控制何时允许用户使用这些操作（根据必须满足的条件）。

在授予权限时，您可以使用 IAM 策略语言和预定义的条件键指定条件。

以下条件键是特定于 Amazon EC2 Auto Scaling 的：

- `autoscaling:ImageId`
- `autoscaling:InstanceType`
- `autoscaling:InstanceTypes`
- `autoscaling:LaunchConfigurationName`
- `autoscaling:LaunchTemplateVersionSpecified`
- `autoscaling:LoadBalancerNames`
- `autoscaling:MaxSize`
- `autoscaling:MinSize`
- `autoscaling:ResourceTag/key`
- `autoscaling:SpotPrice`
- `autoscaling:TargetGroupARNs`
- `autoscaling:VPCZoneIdentifiers`

Important

有关可限制的 API 操作的完整列表、每个操作支持的条件键以及 AWS 范围的条件键，请参阅 IAM 用户指南 中的 [Amazon EC2 Auto Scaling 的操作、资源和条件键](#) 和 [AWS 全局条件上下文键](#)。

预定义的 AWS 托管策略

AWS 创建的托管策略将授予针对常用案例的必要权限。您可以根据您的 IAM 用户所需的访问权限将这些策略附加到这些用户。每个策略授予 Amazon EC2 Auto Scaling 的全部或部分 API 操作的访问权限。

下面是适用于 Amazon EC2 Auto Scaling 的 AWS 托管策略：

- `AutoScalingConsoleFullAccess` — 授予使用 AWS 管理控制台访问 Amazon EC2 Auto Scaling 的完全权限。
- `AutoScalingConsoleReadOnlyAccess` — 授予使用 AWS 管理控制台访问 Amazon EC2 Auto Scaling 的只读权限。
- `AutoScalingFullAccess` — 授予 Amazon EC2 Auto Scaling 的完全访问权限。
- `AutoScalingReadOnlyAccess` — 授予 Amazon EC2 Auto Scaling 的只读访问权限。

您也可以使用 `AmazonEC2FullAccess` 策略授予所有 Amazon EC2 资源和相关服务的完全访问权限。

客户托管策略示例

您可以创建自己的自定义 IAM 策略，以允许或拒绝 IAM 用户或组执行 Amazon EC2 Auto Scaling 操作的权限。您可以将这些自定义策略附加到需要指定权限的 IAM 用户或组。以下示例显示了几个常见使用情形的权限。

示例：限制能够传递哪些服务相关角色（使用 `PassRole`）

如果您的用户需要能够将自定义服务相关角色传递到 Auto Scaling 组，请根据所需访问权限将策略附加到用户或角色。我们建议您将此策略限制为仅针对用户必须访问的服务相关角色。

在您将不同服务相关角色的访问权限提供给不同密钥时，以下示例有助于提升客户托管 CMK 的安全性。根据需求，您可能有三个 CMK 分别供开发团队、QA 团队和财务团队使用。首先，创建具有对所需 CMK 访问权限的服务相关角色，例如，名为 `AWSServiceRoleForAutoScaling_devteamkeyaccess` 的服务相关角色。然后，要授予权限以将该服务相关角色传递到 Auto Scaling 组，请将策略附加到您的 IAM 用户，如下所示。

以下示例中的策略向用户授予权限，传递 `AWSServiceRoleForAutoScaling_devteamkeyaccess` 角色用于创建名称以 `devteam` 开头的任意 Auto Scaling。如果用户尝试指定不同的服务相关角色，则会收到错误。如果用户选择不指定服务相关角色，则改为使用默认 `AWSServiceRoleForAutoScaling` 角色。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:PassRole",
      "Resource": "arn:aws:iam::123456789012:role/aws-service-role/autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling_devteamkeyaccess",
      "Condition": {
        "StringEquals": {
          "iam:PassedToService": [
            "autoscaling.amazonaws.com"
          ]
        },
        "StringLike": {
          "iam:AssociatedResourceARN": [
            "arn:aws:autoscaling:region:123456789012:autoScalingGroup:*:autoScalingGroupName/devteam*"
          ]
        }
      }
    }
  ]
}
```

有关更多信息，请参阅 [适用于 Amazon EC2 Auto Scaling 的服务相关角色 \(p. 143\)](#) 和 [与加密卷结合使用时必需的 CMK 密钥策略 \(p. 145\)](#)。

示例：需要启动模板

以下策略向 IAM 用户授予创建和更新 Auto Scaling 组的权限，前提是他们使用启动模板并指定组用于启动实例的启动模板版本。每个实例在启动期间使用用户指定的启动模板版本。用户可以访问在启动模板中指定的 Amazon EC2 资源。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling:CreateAutoScalingGroup",
        "autoscaling:UpdateAutoScalingGroup"
      ],
      "Resource": "*",
      "Condition": {
        "Bool": {
          "autoscaling:LaunchTemplateVersionSpecified": "true"
        }
      }
    }
  ]
}
```

```
        "Effect": "Allow",
        "Action": [
            "ec2:*"
        ],
        "Resource": "*"
    }
}
```

autoscaling:LaunchTemplateVersionSpecified 条件键接受以下值：

- true - 确保制定启动模板版本。
- false - 确保指定 Latest 或 Default 启动模板版本。
- null - 确保未指定启动模板。

ec2:* 授予调用所有 Amazon EC2 API 操作和访问所有 Amazon EC2 资源的权限。

示例：创建并管理启动配置

以下策略允许用户使用在名称中包含 LaunchConfiguration 字符串的所有 Amazon EC2 Auto Scaling 操作。或者，您还可以明确地列出每项操作，而不是使用通配符。但是，此策略不会自动应用于其名称中包含 LaunchConfiguration 的任何新 Amazon EC2 Auto Scaling 操作。

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": "autoscaling:*LaunchConfiguration*",
    "Resource": "*"
  }]
}
```

以下策略授权用户在实例类型为 t2.micro 且启动配置的名称以 t2micro- 开头时创建启动配置。这些策略可以仅为名称以 t2micro- 开头的 Auto Scaling 组指定启动配置。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "autoscaling:CreateLaunchConfiguration",
      "Resource": [
        "arn:aws:autoscaling:region:123456789012:launchConfiguration:*:launchConfigurationName/t2micro-*"
      ],
      "Condition": {
        "StringEquals": { "autoscaling:InstanceType": "t2.micro" }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling:CreateAutoScalingGroup",
        "autoscaling:UpdateAutoScalingGroup"
      ],
      "Resource": "*",
      "Condition": {
        "StringLikeIfExists": { "autoscaling:LaunchConfigurationName": "t2micro-*" }
      }
    }
  ]
}
```



```
    }  
  }]  
}
```

示例：创建并管理 Auto Scaling 组和扩展策略

以下策略允许用户使用在名称中包含 Scaling 字符串的所有 Amazon EC2 Auto Scaling 操作。

```
{  
  "Version": "2012-10-17",  
  "Statement": [{  
    "Effect": "Allow",  
    "Action": ["autoscaling:*Scaling*"],  
    "Resource": "*" ]  
}]  
}
```

以下策略授权用户使用其名称中包含 Scaling 字符串的所有 Amazon EC2 Auto Scaling 操作，前提是 Auto Scaling 组具有标签 purpose=webserver。由于 Describe 操作不支持资源级权限，因此，您必须在不带条件的单独语句中必须指定它们。

```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Effect": "Allow",  
      "Action": ["autoscaling:*Scaling*"],  
      "Resource": "*",  
      "Condition": {  
        "StringEquals": { "autoscaling:ResourceTag/purpose": "webserver" }  
      }  
    },  
    {  
      "Effect": "Allow",  
      "Action": "autoscaling:Describe*Scaling*",  
      "Resource": "*" ]  
}]  
}
```

以下策略授权用户使用其名称中包含字符串 Scaling 的所有 Amazon EC2 Auto Scaling 操作，前提是它们未指定小于 1 的最小大小或大于 10 的最大大小。由于 Describe 操作不支持资源级权限，因此您必须在不带条件的单独语句中必须指定它们。

```
{  
  "Version": "2012-10-17",  
  "Statement": [  
    {  
      "Effect": "Allow",  
      "Action": ["autoscaling:*Scaling*"],  
      "Resource": "*",  
      "Condition": {  
        "NumericGreaterThanEqualsIfExists": { "autoscaling:MinSize": 1 },  
        "NumericLessThanEqualsIfExists": { "autoscaling:MaxSize": 10 }  
      }  
    },  
    {  
      "Effect": "Allow",  
      "Action": "autoscaling:Describe*Scaling*",  
      "Resource": "*" ]  
}]
```

```
}
```

示例：使用标签控制访问

要授予用户创建 Auto Scaling 组或为该组添加标签（仅当用户指定特定标签时）的权限，请使用 `aws:RequestTag` 条件键。要仅允许特定的标签键，请使用带 `ForAnyValue` 修饰符的 `aws:TagKeys` 条件键。

以下策略需要用户使用标签 `purpose=webserver` 和 `cost-center=cc123` 标记任何 Auto Scaling 组，并且仅允许 `purpose` 和 `cost-center` 标签（不能指定其他标签）。

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "autoscaling:CreateAutoScalingGroup",
      "autoscaling:CreateOrUpdateTags"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "aws:RequestTag/purpose": "webserver",
        "aws:RequestTag/cost-center": "cc123"
      },
      "ForAllValues:StringEquals": { "aws:TagKeys": ["purpose", "cost-center"] }
    }
  }]
}
```

以下策略需要用户在请求中指定包含键 `environment` 的标签。

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "autoscaling:CreateAutoScalingGroup",
      "autoscaling:CreateOrUpdateTags"
    ],
    "Resource": "*",
    "Condition": {
      "StringLike": { "aws:RequestTag/environment": "*" }
    }
  }]
}
```

以下策略需要用户在请求中指定至少一个标签，并且仅允许 `cost-center` 和 `owner` 键。

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "autoscaling:CreateAutoScalingGroup",
      "autoscaling:CreateOrUpdateTags"
    ],
    "Resource": "*",
    "Condition": {
      "ForAnyValue:StringEquals": { "aws:TagKeys": ["cost-center", "owner"] }
    }
  }]
}
```

```
    }  
  }]  
}
```

以下策略授权用户访问具有标签 `allowed=true` 的 Auto Scaling 组，并允许它们仅应用标签 `environment=test`。由于启动配置不支持标签并且 `Describe` 操作不支持资源级权限，因此您必须在不带条件的单独语句中指定它们。

```
{  
  "Version": "2012-10-17",  
  "Statement": [{  
    "Effect": "Allow",  
    "Action": "autoscaling:*Scaling*",  
    "Resource": "*",  
    "Condition": {  
      "StringEquals": { "autoscaling:ResourceTag/allowed": "true" },  
      "StringEqualsIfExists": { "aws:RequestTag/environment": "test" },  
      "ForAllValues:StringEquals": { "aws:TagKeys": "environment" }  
    }  
  },  
  {  
    "Effect": "Allow",  
    "Action": [  
      "autoscaling:*LaunchConfiguration*",  
      "autoscaling:Describe*"  
    ],  
    "Resource": "*"  
  }]  
}
```

示例：更改 Auto Scaling 组的容量

以下策略授权用户使用 `SetDesiredCapacity` 操作更改任何 Auto Scaling 组的容量。

```
{  
  "Version": "2012-10-17",  
  "Statement": [{  
    "Effect": "Allow",  
    "Action": "autoscaling:SetDesiredCapacity",  
    "Resource": "*"  
  }]  
}
```

以下策略授权用户使用 `SetDesiredCapacity` 操作更改指定的 Auto Scaling 组的容量。包括 UUID 可确保将访问权授予特定的 Auto Scaling 组。新组的 UUID 与删除的同名组的 UUID 不同。

```
{  
  "Version": "2012-10-17",  
  "Statement": [{  
    "Effect": "Allow",  
    "Action": "autoscaling:SetDesiredCapacity",  
    "Resource": [  
  
      "arn:aws:autoscaling:region:123456789012:autoScalingGroup:7fe02b8e-7442-4c9e-8c8e-85fa99e9b5d9:autoScalingGroup-1",  
      "arn:aws:autoscaling:region:123456789012:autoScalingGroup:9d8e8ea4-22e1-44c7-a14d-520f8518c2b9:autoScalingGroupName/group-2",  
      "arn:aws:autoscaling:region:123456789012:autoScalingGroup:60d6b363-ae8b-467c-947f-f1d308935521:autoScalingGroupName/group-3"  
    ]  
  }]  
}
```

```
    }}  
  }
```

以下策略授权用户使用 `SetDesiredCapacity` 操作更改名称以 `group-` 开头的任何 Auto Scaling 组的容量。

```
{  
  "Version": "2012-10-17",  
  "Statement": [{  
    "Effect": "Allow",  
    "Action": "autoscaling:SetDesiredCapacity",  
    "Resource": [  
      "arn:aws:autoscaling:region:123456789012:autoScalingGroup:*:autoScalingGroupName/  
group-*"  
    ]  
  }]  
}
```

适用于 Amazon EC2 Auto Scaling 的服务相关角色

Amazon EC2 Auto Scaling 使用服务相关角色获取代表您调用其他 AWS 服务所需的权限。服务相关角色是一种独特类型的 IAM 角色，它与 AWS 服务直接关联。

服务相关角色提供了一种将权限委托给 AWS 服务的安全方式，因为只有相关服务才能代入服务相关角色。有关更多信息，请参阅 IAM 用户指南 中的 [使用服务相关角色](#)。

AWSServiceRoleForAutoScaling 授予的权限

Amazon EC2 Auto Scaling 使用名为 `AWSServiceRoleForAutoScaling` 的服务相关角色代表您管理 Auto Scaling 组。`AWSServiceRoleForAutoScaling` 角色自动分配给 Auto Scaling 组，除非您指定不同的服务相关角色。

为 `AWSServiceRoleForAutoScaling` 角色预定义了代表您进行以下调用的权限：

- `ec2:AttachClassicLinkVpc`
- `ec2:CancelSpotInstanceRequests`
- `ec2:CreateTags`
- `ec2>DeleteTags`
- `ec2:Describe*`
- `ec2:DetachClassicLinkVpc`
- `ec2:ModifyInstanceAttribute`
- `ec2:RequestSpotInstances`
- `ec2:RunInstances`
- `ec2:TerminateInstances`
- `elasticloadbalancing:Register*`
- `elasticloadbalancing:Deregister*`
- `elasticloadbalancing:Describe*`
- `cloudwatch>DeleteAlarms`
- `cloudwatch:DescribeAlarms`
- `cloudwatch:PutMetricAlarm`
- `sns:Publish`

此角色信任 `autoscaling.amazonaws.com` 服务来代入它。

您必须配置权限以允许 IAM 实体 (如用户、组或角色) 创建、编辑或删除服务相关角色。有关更多信息, 请参阅 IAM 用户指南 中的 [使用服务相关角色](#)。

创建服务相关角色 (自动)

大多数情况下, 您无需手动创建服务相关角色。在首次创建 Auto Scaling 组但未指定其他服务相关角色时, Amazon EC2 Auto Scaling 将为您创建 `AWSServiceRoleForAutoScaling` 服务相关角色。

如果在 2018 年 3 月 (此时, Amazon EC2 Auto Scaling 开始支持服务相关角色) 之前创建 Auto Scaling 组, Amazon EC2 Auto Scaling 将在您的 AWS 账户中创建 `AWSServiceRoleForAutoScaling` 角色。有关更多信息, 请参阅 IAM 用户指南 中的 [我的 AWS 账户中出现新角色](#)。

Important

确保您已启用允许 IAM 实体 (如用户、组或角色) 创建服务相关角色的 IAM 权限。否则, 自动创建操作将失败。有关更多信息, 请参阅 IAM 用户指南 中的 [服务相关角色权限](#) 或本指南中有关 [所需的用户权限 \(p. 134\)](#) 的信息。

创建服务相关角色 (手动)

或者, 在创建 Auto Scaling 组时, 您也可以创建然后指定自己的服务相关角色 (含自定义后缀)。如果您必须为不同的服务相关角色授予对不同的客户主密钥 (KMS) 的访问权限, 这会很有用。有关更多信息, 请参阅 [与加密卷结合使用时必需的 CMK 密钥策略 \(p. 145\)](#)。您还可以记下正在使用的服务相关角色, 以便跟踪哪个 Auto Scaling 组在您的 CloudTrail 日志中进行了 API 调用。有关将 IAM 用户或角色的访问限制到特定的自定义服务相关角色的策略, 请参阅 [示例: 限制能够传递哪些服务相关角色 \(使用 PassRole\) \(p. 137\)](#)。

您可以使用 IAM 控制台、AWS CLI 或 IAM API 创建服务相关角色。有关更多信息, 请参阅 IAM 用户指南 中的 [创建服务相关角色](#)。

例如, 使用以下 `create-service-linked-role` CLI 命令为 Amazon EC2 Auto Scaling 创建一个名为 `AWSServiceRoleForAutoScaling_`*suffix* 的服务相关角色。该后缀可帮助您确定角色的用途。

```
aws iam create-service-linked-role --aws-service-name autoscaling.amazonaws.com --custom-suffix suffix
```

此命令的输出包含服务相关角色的 ARN, 您可以用来向服务相关角色授予对 CMK 的访问权限。

```
{
  "Role": {
    "RoleId": "ABCDEF0123456789ABCDEF",
    "CreateDate": "2018-08-30T21:59:18Z",
    "RoleName": "AWSServiceRoleForAutoScaling_suffix",
    "Arn": "arn:aws:iam::123456789012:role/aws-service-role/autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling_suffix",
    "Path": "/aws-service-role/autoscaling.amazonaws.com/",
    "AssumeRolePolicyDocument": {
      "Version": "2012-10-17",
      "Statement": [
        {
          "Action": [
            "sts:AssumeRole"
          ],
          "Principal": {
            "Service": [
              "autoscaling.amazonaws.com"
            ]
          },
          "Effect": "Allow"
        }
      ]
    }
  }
}
```

```
}  
    ]  
  }  
}
```

编辑服务相关角色

您可以编辑您为了与 Auto Scaling 组结合使用而创建的任何服务相关角色。有关更多信息，请参阅 IAM 用户指南 中的[编辑服务相关角色](#)。

对于 Amazon EC2 Auto Scaling 创建的 `AWSServiceRoleForAutoScaling` 角色，您只能编辑其描述，而不能编辑其权限。

删除服务相关角色

如果您不再需要使用某个 Auto Scaling 组，我们建议您删除其服务相关角色。只有在首先删除相关 AWS 资源后，才能删除服务相关角色。如果某个服务相关角色与多个 Auto Scaling 组结合使用，则必须删除使用该服务相关角色的所有 Auto Scaling 组，然后才能删除该服务相关角色。这会保护您的 Auto Scaling 组，因为您无法意外删除这些资源的管理权限。有关更多信息，请参阅[删除 Auto Scaling 基础设施](#) (p. 61)。

您可以使用 IAM 来删除默认服务相关角色或您创建的服务相关角色。有关更多信息，请参阅 IAM 用户指南 中的[删除服务相关角色](#)。

如果删除 `AWSServiceRoleForAutoScaling` 服务相关角色，在您创建 Auto Scaling 组但未指定其他服务相关角色时，Amazon EC2 Auto Scaling 将再次创建该角色。

Amazon EC2 Auto Scaling 服务相关角色支持的区域

Amazon EC2 Auto Scaling 支持在提供该服务的所有 AWS 区域中使用服务相关角色。有关更多信息，请参阅[AWS 区域和终端节点](#)。

与加密卷结合使用时必需的 CMK 密钥策略

如果您指定客户管理的客户主密钥 (CMK) 以实现 Amazon EBS 加密，您必须授予对 CMK 的相应服务角色的访问权限，以便 Amazon EC2 Auto Scaling 可以代表您启动实例。为此，您必须在创建 CMK 时或稍后修改 CMK 的密钥策略。

Note

Amazon EC2 Auto Scaling 已授权仅使用用于 EBS 加密的 [AWS 托管 CMK](#)。此 CMK 对于您的账户和使用该账户的 AWS 区域而言是唯一的。AWS 托管 CMK 用于加密，除非您指定[客户托管 CMK](#)。有关创建客户托管 CMK 的信息，请参阅 AWS Key Management Service Developer Guide 中的[创建密钥](#)。

示例：允许访问 CMK 的 CMK 密钥策略部分

以下两个策略部分为名为 `AWSServiceRoleForAutoScaling` 的服务相关角色授予使用客户托管 CMK 的权限。

```
{  
  "Sid": "Allow use of the CMK",  
  "Effect": "Allow",  
  "Principal": {  
    "AWS": [
```

```
        "arn:aws:iam::123456789012:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling"
    ],
    "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
    ],
    "Resource": "*"
}
```

```
{
  "Sid": "Allow attachment of persistent resources",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::123456789012:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling"
    ]
  },
  "Action": [
    "kms:CreateGrant"
  ],
  "Resource": "*",
  "Condition": {
    "Bool": {
      "kms:GrantIsForAWSResource": true
    }
  }
}
```

向 CMK 策略添加新部分时，不要更改策略中任何已存在的部分。有关编辑密钥策略的更多信息，请参阅 AWS Key Management Service Developer Guide 中的[在 AWS KMS 中使用密钥策略](#)。

Important

密钥策略的 Principal 元素是服务相关角色的 Amazon 资源名称 (ARN)。在启动按需实例时，使用适用于 Amazon EC2 Auto Scaling 的服务相关角色的 ARN：AWSServiceRoleForAutoScaling。在启动 Spot 实例时，如果使用了启动配置，则使用 AWSServiceRoleForEC2Spot 角色的 ARN，如果使用了启动模板，则使用适用于 Amazon EC2 Auto Scaling 的服务相关角色。

示例：允许跨账户访问 CMK 的 CMK 密钥策略部分

如果您的客户托管 CMK 与 Auto Scaling 组位于不同的账户中，您必须组合使用授权和密钥策略以允许访问 CMK。有关更多信息，请参阅 AWS Key Management Service Developer Guide 中的[使用授权](#)。

首先，将以下两部分添加到 CMK 的密钥策略中，以便您可以将 CMK 与外部账户（根用户）一起使用。

```
{
  "Sid": "Allow use of the key",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::111122223333:root"
    ]
  },
  "Action": [
    "kms:Encrypt",

```



```
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey"
  ],
  "Resource": "*"
}
```

```
{
  "Sid": "Allow attachment of persistent resources",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::111122223333:root"
    ]
  },
  "Action": [
    "kms:CreateGrant"
  ],
  "Resource": "*"
}
```

然后，从外部账户创建一个授权，将相关权限委派给相应的服务相关角色。该授权的 Grantee Principal 元素是相应的服务相关角色的 ARN。密钥 ID 是您的账户中 CMK 的 ARN。以下是[创建授权 CLI 命令示例](#)，该命令授予名为 `AWSServiceRoleForAutoScaling` 的服务相关角色使用 CMK 权限：

```
aws kms create-grant \
  --region us-west-2 \
  --key-id arn:aws:kms:us-west-2:444455556666:key/1a2b3c4d-5e6f-1a2b-3c4d-5e6f1a2b3c4d \
  --grantee-principal arn:aws:iam::111122223333:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling \
  --operations "Encrypt" "Decrypt" "ReEncryptFrom" "ReEncryptTo" "GenerateDataKey"
"GenerateDataKeyWithoutPlaintext" "DescribeKey" "CreateGrant"
```

如果您在配置启动具有加密卷的实例所需的客户托管 CMK 的跨账户访问时遇到任何问题，请参阅[故障排除部分](#) (p. 153)。

有关更多信息，请参阅 Amazon EC2 用户指南（适用于 Linux 实例）中的 [Amazon EBS Encryption](#) 和 [AWS Key Management Service Developer Guide](#)。

在 Amazon EC2 实例上运行的应用程序的 IAM 角色

在 Amazon EC2 实例上运行的应用程序需要凭证以访问其他 AWS 服务。要以安全的方式提供这些凭证，请使用 IAM 角色。角色可提供临时权限，以供应用程序在访问其他 AWS 资源时使用。角色的许可将确定允许访问资源的应用程序。

有关更多信息，请参阅 IAM 用户指南 中的[使用 IAM 角色向在 Amazon EC2 实例上运行的应用程序授予权限](#)。

对于 Auto Scaling 组中的实例，您必须创建启动配置或模板，并选择要与实例关联的实例配置文件。实例配置文件是 IAM 角色的容器，在实例启动时允许 Amazon EC2 将 IAM 角色传递到实例。首先，创建具有访问 AWS 资源所需的所有权限的 IAM 角色。然后，创建实例配置文件并将该角色分配给它。有关更多信息，请参阅 IAM 用户指南 中的[使用实例配置文件](#)。

Note

当您使用 IAM 控制台为 Amazon EC2 创建角色时，控制台会指导您完成创建角色的步骤，并自动创建与 IAM 角色同名的实例配置文件。

先决条件

创建运行在 Amazon EC2 上的应用程序可代入的 IAM 角色。选择适当的权限，以便随后向应用程序提供可以访问所需资源的角色。

创建 IAM 角色

1. 通过以下网址打开 IAM 控制台：<https://console.aws.amazon.com/iam/>。
2. 在导航窗格中，选择 Roles 和 Create new role。
3. 在 Select role type 页面上，选择 EC2 和 EC2 使用案例。选择 Next: Permissions (下一步: 权限)。
4. 在附加权限策略页面上，选择包含所需权限的 AWS 托管策略。
5. 在 Review 页面上，为角色键入一个名称，然后选择 Create role。

创建启动配置

在使用 AWS 管理控制台创建启动配置时，在 Configure Details (配置详细信息) 页上，从 IAM role (IAM 角色) 中选择角色。有关更多信息，请参阅 [创建启动配置 \(p. 28\)](#)。

从 AWS CLI 中使用 `create-launch-configuration` 命令创建启动配置时，请指定实例配置文件名称，如下所示：

```
aws autoscaling create-launch-configuration --launch-configuration-name my-lc-with-instance-profile \
--image-id ami-01e24be29428c15b2 --instance-type t2.micro \
--iam-instance-profile my-instance-profile
```

创建启动模板

当您使用 AWS 管理控制台创建启动模板时，在高级详细信息部分中，从 IAM 实例配置文件中选择角色。有关更多信息，请参阅 [为 Auto Scaling 组创建启动模板 \(p. 22\)](#)。

从 AWS CLI 中使用 `create-launch-template` 命令创建启动模板时，请指定实例配置文件名称，如下所示：

```
aws ec2 create-launch-template --launch-template-name my-lt-with-instance-profile --
version-description version1 \
--launch-template-data
'{"ImageId":"ami-01e24be29428c15b2","InstanceType":"t2.micro","IamInstanceProfile":
{"Name":"my-instance-profile"}}'
```

排查 Amazon EC2 Auto Scaling 问题

Amazon EC2 Auto Scaling 提供特定的描述性错误消息来帮助您排查问题。可以从扩展活动的描述中发现错误消息。

目录

- [检索错误消息 \(p. 149\)](#)
- [Amazon EC2 Auto Scaling 问题排查：EC2 实例启动失败 \(p. 151\)](#)
- [Amazon EC2 Auto Scaling 问题排查：AMI 问题 \(p. 154\)](#)
- [Amazon EC2 Auto Scaling 问题排查：负载均衡器问题 \(p. 155\)](#)
- [Auto Scaling 问题排查：容量限制 \(p. 156\)](#)

检索错误消息

要从扩展活动描述中检索错误消息，请使用 `describe-scaling-activities` 命令，如下所示：

```
aws autoscaling describe-scaling-activities --auto-scaling-group-name my-asg
```

在下面的示例响应中，`StatusCode` 包含活动的当前状态，`StatusMessage` 包含错误消息：

```
{
  "Activities": [
    {
      "Description": "Launching a new EC2 instance: i-4ba0837f",
      "AutoScalingGroupName": "my-asg",
      "ActivityId": "f9f2d65b-f1f2-43e7-b46d-d86756459699",
      "Details": "{\"Availability Zone\":\"us-west-2c\"}",
      "StartTime": "2013-08-19T20:53:29.930Z",
      "Progress": 100,
      "EndTime": "2013-08-19T20:54:02Z",
      "Cause": "At 2013-08-19T20:53:25Z a user request created an
AutoScalingGroup...",
      "StatusCode": "Failed",
      "StatusMessage": "The image id 'ami-4edb0327' does not exist. Launching EC2
instance failed."
    }
  ]
}
```

下表列出了错误消息的类型，并提供问题排查资源的链接，供您在排查问题时使用。

EC2 实例启动失败

问题	错误消息
Auto Scaling 组	AutoScalingGroup <Auto Scaling 组名称> 未找到。 (p. 152)
可用区	不再支持请求的可用区。请重新尝试请求... (p. 152)
AWS 账户	您没有订阅此项服务。请参阅 https://aws.amazon.com/。 (p. 152)

问题	错误消息
块储存设备映射	设备名称上载无效。启动 EC2 实例失败。(p. 152)
块储存设备映射	用于参数 virtualName 的值 (<与实例存储设备相关联的名称>) 无效... (p. 153)
块储存设备映射	实例存储 AMI 不支持 EBS 块储存设备映射。(p. 153)
实例类型和可用区	请求的实例类型 (<实例类型>) 在请求的可用区 (<实例可用区>) 中不受支持... (p. 152)
密钥对	密钥对 <与您的 EC2 实例相关联的密钥对> 不存在。启动 EC2 实例失败。(p. 151)
启动配置	当前不支持请求的配置。(p. 152)
置放群组	置放群组可能无法与类型为“m1.large”的实例一起使用。启动 EC2 实例失败。(p. 153)
安全组	安全组 <该安全组的名称> 不存在。启动 EC2 实例失败。(p. 151)
服务相关角色	Client.InternalError : 客户端在启动时出错。(p. 153)

AMI 问题

问题	错误消息
AMI ID	AMI ID <您的 AMI 的 ID> 不存在。启动 EC2 实例失败。(p. 154)
AMI ID	AMI <AMI ID> 正在等待，无法运行。启动 EC2 实例失败。(p. 155)
AMI ID	用于参数 virtualName 的值 (<ami ID>) 无效。(p. 155)
架构不匹配	请求的实例类型的架构 (i386) 与用于 ami-6622f00f (x86_64) 的清单中的架构不匹配。启动 EC2 实例失败。(p. 155)

负载均衡器问题

问题	错误消息
找不到负载均衡器	找不到负载均衡器 <您的启动环境>。验证负载均衡器配置失败。(p. 156)
VPC 中的实例	EC2 实例 <实例 ID> 不在 VPC 内。更新负载均衡器配置失败。(p. 156)
没有活动的负载均衡器	名为 <负载均衡器名称> 的活动负载均衡器不存在。更新负载均衡器配置失败。(p. 156)
安全令牌	请求中包含的安全令牌无效。验证负载均衡器配置失败。(p. 156)

容量限制

问题	错误消息
容量限制	已运行 <实例数量> 个实例。启动 EC2 实例失败。(p. 157)
可用区中容量不足	您请求的可用区 (<请求的可用区>) 中当前没有足够的 <实例类型> 容量... (p. 157)

Amazon EC2 Auto Scaling 问题排查：EC2 实例启动失败

此页提供有关未能启动的 EC2 实例的信息、可能的原因，以及可用来解决问题的步骤。

要检索错误消息，请参阅[检索错误消息](#) (p. 149)。

EC2 实例启动失败时，您可能会收到一条或多条以下错误消息：

错误消息

- 安全组 <该安全组的名称> 不存在。启动 EC2 实例失败。 (p. 151)
- 密钥对 <与您的 EC2 实例相关联的密钥对> 不存在。启动 EC2 实例失败。 (p. 151)
- 当前不支持请求的配置。 (p. 152)
- AutoScalingGroup <Auto Scaling 组名称> 未找到。 (p. 152)
- 不再支持请求的可用区。请重新尝试请求... (p. 152)
- 请求的实例类型 (<实例类型>) 在请求的可用区 (<实例可用区>) 中不受支持... (p. 152)
- 您没有订阅此项服务。请参阅 <https://aws.amazon.com/>。 (p. 152)
- 设备名称上载无效。启动 EC2 实例失败。 (p. 152)
- 用于参数 virtualName 的值 (<与实例存储设备相关联的名称>) 无效... (p. 153)
- 实例存储 AMI 不支持 EBS 块储存设备映射。 (p. 153)
- 置放群组可能无法与类型为“m1.large”的实例一起使用。启动 EC2 实例失败。 (p. 153)
- Client.InternalError：客户端在启动时出错。 (p. 153)

安全组 <该安全组的名称> 不存在。启动 EC2 实例失败。

- 原因：可能已删除启动配置中指定的安全组。
- 解决方案：
 1. 使用 [describe-security-groups](#) 命令获取与您的账户关联的安全组列表。
 2. 从该列表中选择要使用的安全组。要创建安全组，请使用 [create-security-group](#) 命令。
 3. 创建新的启动配置。
 4. 使用 [update-auto-scaling-group](#) 命令通过新的启动配置更新 Auto Scaling 组。

密钥对 <与您的 EC2 实例相关联的密钥对> 不存在。启动 EC2 实例失败。

- 原因：可能已删除启动实例时使用的密钥对。
- 解决方案：
 1. 使用 [describe-key-pairs](#) 命令获取可用的密钥对列表。
 2. 从该列表中选择要使用的密钥对。要创建密钥对，请使用 [create-key-pair](#) 命令。
 3. 创建新的启动配置。
 4. 使用 [update-auto-scaling-group](#) 命令通过新的启动配置更新 Auto Scaling 组。

当前不支持请求的配置。

- 原因：启动配置中的某些选项当前可能不受支持。
- 解决方案：
 1. 创建新的启动配置。
 2. 使用 `update-auto-scaling-group` 命令通过新的启动配置更新 Auto Scaling 组。

AutoScalingGroup <Auto Scaling 组名称> 未找到。

- 原因：该 Auto Scaling 组可能已删除。
- 解决方案：创建新 Auto Scaling 组。

不再支持请求的可用区。请重新尝试请求...

- Error Message：不再支持请求的可用区。请通过不指定可用区或选择 <提供的可用区的列表> 来重新尝试请求。启动 EC2 实例失败。
- 原因：当前可能无法使用与 Auto Scaling 组相关联的可用区。
- 解决方案：采用错误消息中的建议更新 Auto Scaling 组。

请求的实例类型 (<实例类型>) 在请求的可用区 (<实例可用区>) 中不受支持...

- 错误消息：您请求的实例类型 (<实例类型>) 在请求的可用区 (<实例可用区>) 中不受支持。请通过不指定可用区或选择 <支持该实例类型的可用区列表> 来重新尝试请求。启动 EC2 实例失败。
- 原因：当前可能无法在 Auto Scaling 组中指定的可用区内使用与启动配置相关联的实例类型。
- 解决方案：采用错误消息中的建议更新 Auto Scaling 组。

您没有订阅此项服务。请参阅 <https://aws.amazon.com/>。

- 原因：您的 AWS 账户可能已过期。
- 解决方案：转到 <https://aws.amazon.com/>，然后选择立即注册以开立新账户。

设备名称上载无效。启动 EC2 实例失败。

- 原因：启动配置中的块储存设备映射所包含的块储存设备名称可能无法使用或目前不受支持。
- 解决方案：
 1. 使用 `describe-volumes` 命令查看如何向实例公开卷。
 2. 使用卷描述中列出的设备名称创建新的启动配置。
 3. 使用 `update-auto-scaling-group` 命令通过新的启动配置更新 Auto Scaling 组。

用于参数 virtualName 的值 (<与实例存储设备相关联的名称>) 无效...

- 错误消息：virtualName 参数的值 (<与实例存储设备相关联的名称>) 无效。预期格式：“ephemeralNUMBER”。启动 EC2 实例失败。
- 原因：与块存储设备相关联的虚拟名称的指定格式不正确。
- 解决方案：
 1. 通过在 virtualName 参数中指定设备名称创建新的启动配置。有关设备名称格式的信息，请参阅 Amazon EC2 用户指南 (适用于 Linux 实例) 中的 [实例存储设备名称](#)。
 2. 使用 [update-auto-scaling-group](#) 命令通过新的启动配置更新 Auto Scaling 组。

实例存储 AMI 不支持 EBS 块储存设备映射。

- 原因：启动配置中指定的块储存设备映射在您的实例上不受支持。
- 解决方案：
 1. 使用实例类型支持的块储存设备映射来创建新的启动配置。有关更多信息，请参阅 Amazon EC2 用户指南 (适用于 Linux 实例) 中的 [块储存设备映射](#)。
 2. 使用 [update-auto-scaling-group](#) 命令通过新的启动配置更新 Auto Scaling 组。

置放群组可能无法与类型为“m1.large”的实例一起使用。启动 EC2 实例失败。

- 原因：您的集群置放群组包含无效实例类型。
- 解决方案：
 1. 有关置放群组所支持的有效实例类型的信息，请参阅 Amazon EC2 用户指南 (适用于 Linux 实例) 中的 [置放群组](#)。
 2. 按照 [置放群组](#) 中的详细说明创建新的置放群组。
 3. 或者，也可以使用受支持的实例类型创建新的启动配置。
 4. 使用 [update-auto-scaling-group](#) 命令通过新的置放群组或启动配置更新 Auto Scaling 组。

Client.InternalError：客户端在启动时出错。

- 原因：在 Auto Scaling 组尝试启动具有加密 EBS 卷的实例，但 [服务相关角色](#) (p. 143) 没有权限访问加密该卷所用的 [客户管理的 CMK](#) 时，就会出现此错误。
- 解决方案：需要进行额外的设置以允许 Auto Scaling 组启动实例。下表总结了解决错误的步骤。有关更多信息，请参阅 <https://forums.aws.amazon.com/thread.jspa?threadID=277523>。

场景	后续步骤
方案 1: CMK 和 Auto Scaling 组属于同一 AWS 账户	允许服务相关角色使用 CMK，如下所示： <ol style="list-style-type: none">1. 确定将哪个服务相关角色用于此 Auto Scaling 组。2. 更新 CMK 上的密钥策略并允许服务相关角色使用 CMK。3. 更新 Auto Scaling 组来使用服务相关角色。

场景	后续步骤
方案 2: CMK 和 Auto Scaling 组属于不同 AWS 账户	<p>有两种可能的解决方案：</p> <p>解决方案 1：使用与 Auto Scaling 组相同的 AWS 账户中的 CMK</p> <ol style="list-style-type: none">1. 使用属于与 Auto Scaling 组相同的账户中的另一个 CMK，复制并重新加密快照。2. 允许服务相关角色使用新 CMK。请参阅方案 1 的步骤。 <p>解决方案 2：继续使用与 Auto Scaling 组不同的 AWS 账户中的 CMK</p> <ol style="list-style-type: none">1. 确定将哪个服务相关角色用于此 Auto Scaling 组。2. 允许 Auto Scaling 组账户（根用户）访问 CMK。3. 在 Auto Scaling 组账户中定义可以创建授权的 IAM 用户或角色。4. 使用服务相关角色作为被授权委托人，创建对 CMK 的授权。5. 更新 Auto Scaling 组来使用服务相关角色。

Amazon EC2 Auto Scaling 问题排查：AMI 问题

本页提供与 AMI 相关联的问题的有关信息、可能原因，以及可用来解决这些问题的步骤。

要检索错误消息，请参阅[检索错误消息](#) (p. 149)。

当 AMI 的问题导致 EC2 实例启动失败时，您可能会收到一条或多条以下错误消息：

错误消息

- AMI ID <您的 AMI 的 ID> 不存在。启动 EC2 实例失败。 (p. 154)
- AMI <AMI ID> 正在等待，无法运行。启动 EC2 实例失败。 (p. 155)
- 用于参数 virtualName 的值 (<ami ID>) 无效。 (p. 155)
- 请求的实例类型的架构 (i386) 与用于 ami-6622f00f (x86_64) 的清单中的架构不匹配。启动 EC2 实例失败。 (p. 155)

AMI ID <您的 AMI 的 ID> 不存在。启动 EC2 实例失败。

- 原因：创建启动配置后，可能已删除 AMI。
- 解决方案：
 1. 使用有效 AMI 创建新的启动配置。
 2. 使用 [update-auto-scaling-group](#) 命令通过新的启动配置更新 Auto Scaling 组。

AMI <AMI ID> 正在等待，无法运行。启动 EC2 实例失败。

- 原因：您可能刚创建 AMI（通过获取运行实例的快照或任何其他方式），它可能还无法使用。
- 解决方案：必须等待您的 AMI 可用后，才能创建启动配置。

用于参数 virtualName 的值 (<ami ID>) 无效。

- 原因：不正确的值。virtualName 参数表示与设备相关联的虚拟名称。
- 解决方案：
 1. 通过指定使用 virtualName 参数的实例的虚拟设备名称来创建新的启动配置。
 2. 使用 `update-auto-scaling-group` 命令通过新的启动配置更新 Auto Scaling 组。

请求的实例类型的架构 (i386) 与用于 ami-6622f00f (x86_64) 的清单中的架构不匹配。启动 EC2 实例失败。

- 原因：启动配置中提到的 InstanceType 的架构与映像架构不匹配。
- 解决方案：
 1. 使用与请求的实例类型的架构相匹配的 AMI 架构来创建新的启动配置。
 2. 使用 `update-auto-scaling-group` 命令通过新的启动配置更新 Auto Scaling 组。

Amazon EC2 Auto Scaling 问题排查：负载均衡器问题

本页提供与 Auto Scaling 组相关联的负载均衡器所导致的问题有关的信息、可能原因，以及可用来解决这些问题的步骤。

要检索错误消息，请参阅[检索错误消息 \(p. 149\)](#)。

如果与 Auto Scaling 组关联的负载均衡器导致 EC2 实例启动失败，您可能会收到一条或多条以下错误消息：

错误消息

- 找不到负载均衡器 <您的启动环境>。验证负载均衡器配置失败。 (p. 156)
- 名为 <负载均衡器名称> 的活动负载均衡器不存在。更新负载均衡器配置失败。 (p. 156)
- EC2 实例 <实例 ID> 不在 VPC 内。更新负载均衡器配置失败。 (p. 156)
- EC2 实例 <实例 ID> 在 VPC 中。更新负载均衡器配置失败。 (p. 156)
- 请求中包含的安全令牌无效。验证负载均衡器配置失败。 (p. 156)

找不到负载均衡器 <您的启动环境>。验证负载均衡器配置失败。

- 原因 1：已删除负载均衡器。
- 解决方案 1：
 1. 检查负载均衡器是否仍然存在。您可以使用 `describe-load-balancers` 命令。
 2. 如果看到响应中列出了负载均衡器，请参阅原因 2。
 3. 如果没有看到响应中列出负载均衡器，可以创建新的负载均衡器，然后创建新的 Auto Scaling 组，或者可以创建无负载均衡器的新的 Auto Scaling 组。
- 原因 2：创建 Auto Scaling 组时，未采用正确顺序指定负载均衡器名称。
- 解决方案 2：创建新的 Auto Scaling 组，并在最后指定负载均衡器名称。

名为 <负载均衡器名称> 的活动负载均衡器不存在。更新负载均衡器配置失败。

- 原因：可能已删除指定的负载均衡器。
- 解决方案：可以创建新的负载均衡器，然后创建新的 Auto Scaling 组，也可以创建无负载均衡器的新 Auto Scaling 组。

EC2 实例 <实例 ID> 不在 VPC 内。更新负载均衡器配置失败。

- 原因：VPC 中不存在指定的实例。
- 解决方案：可以删除与实例相关联的负载均衡器，或者创建新的 Auto Scaling 组。

EC2 实例 <实例 ID> 在 VPC 中。更新负载均衡器配置失败。

- 原因：负载均衡器在 EC2-Classic 中，但 Auto Scaling 组在 VPC 中。
- 解决方案：确保负载均衡器和 Auto Scaling 组在同一网络（EC2-Classic 或 VPC）中。

请求中包含的安全令牌无效。验证负载均衡器配置失败。

- 原因：您的 AWS 账户可能已过期。
- 解决方案：检查 AWS 账户是否有效。转到 <https://aws.amazon.com/>，然后选择立即注册以开立新账户。

Auto Scaling 问题排查：容量限制

本页提供有关 Auto Scaling 组的容量限制问题的信息、可能原因，以及可用来解决这些问题的步骤。有关 Amazon EC2 Auto Scaling 限制的更多信息，请参阅 [Amazon EC2 Auto Scaling 限制 \(p. 8\)](#)。

要检索错误消息，请参阅[检索错误消息](#) (p. 149)。

如果 Auto Scaling 组的容量限制问题导致 EC2 实例启动失败，您可能会收到一条或多条以下错误消息。

错误消息

- 您请求的可用区 (<请求的可用区>) 中当前没有足够的 <实例类型> 容量... (p. 157)
- 已运行 <实例数量> 个实例。启动 EC2 实例失败。 (p. 157)

您请求的可用区 (<请求的可用区>) 中当前没有足够的 <实例类型> 容量...

- 错误消息：您请求的可用区 (<请求的可用区>) 中当前没有足够的 <实例类型> 容量。我们的系统将预置额外的容量。您当前可以通过不在请求中指定可用区，或者选择 <当前支持该实例类型的可用区列表> 来获取 <实例类型> 容量。启动 EC2 实例失败。
- 原因：现在，Auto Scaling 不能在请求的可用区中支持您的实例类型。
- 解决方案：
 1. 按照错误消息中的建议创建新的启动配置。
 2. 使用 `update-auto-scaling-group` 命令通过新的启动配置更新 Auto Scaling 组。

已运行 <实例数量> 个实例。启动 EC2 实例失败。

- 原因：Auto Scaling 组已达到 `DesiredCapacity` 参数设置的限制。
- 解决方案：
 - 使用 `update-auto-scaling-group` 命令为 `--desired-capacity` 参数提供新的值以更新 Auto Scaling 组。
 - 如果您已达到 EC2 实例的数量限制，可以请求增加。有关更多信息，请参阅 [AWS 服务限制](#)。

Auto Scaling 资源

下列相关资源在您使用此服务的过程中会有所帮助。

- [Amazon EC2 Auto Scaling](#) – 提供 Amazon EC2 Auto Scaling 相关信息的主要网页。
- [Amazon EC2 技术常见问题](#) – 客户提出的 Amazon EC2 和 Amazon EC2 Auto Scaling 问题的答案。
- [Amazon EC2 论坛](#) – 从社区获取帮助。
- [AWS Auto Scaling 用户指南](#) – 使用 AWS Auto Scaling 控制台，您可以更轻松地使用多个服务的扩展功能。借助 AWS Auto Scaling，您还可以简化为 Auto Scaling 组定义动态扩展策略的流程，并使用预测扩展功能在提前预测到流量变化时扩展您的 Amazon EC2 容量。

以下附加资源可帮助您了解有关 AWS 的更多信息。

- [课程和研讨会](#) – 指向基于角色的专业课程和自主进度动手实验室的链接，这些课程和实验室旨在帮助您增强 AWS 技能并获得实践经验。
- [AWS 开发人员工具](#) – 指向开发人员工具、软件开发工具包、IDE 工具包和命令行工具的链接，这些资源用于开发和管理 AWS 应用程序。
- [AWS 白皮书](#) – 指向 AWS 技术白皮书的完整列表的链接，这些资料涵盖了架构、安全性、经济性等主题，由 AWS 解决方案架构师或其他技术专家编写。
- [AWS Support 中心](#) - 用于创建和管理 AWS Support 案例的中心。还包括指向其他有用资源的链接，如论坛、技术常见问题、服务运行状况和 AWS Trusted Advisor。
- [AWS Support](#) - 提供有关 AWS Support 信息的主要网页，是一种一对一的快速响应支持渠道，可帮助您在云中构建和运行应用程序。
- [联系我们](#) - 查询有关 AWS 账单、账户、事件、滥用和其他问题的中央联系点。
- [AWS 网站条款](#) - 有关我们的版权和商标、您的账户、许可、网站访问和其他主题的详细信息。

文档历史记录

下表介绍了自 2018 年 7 月以来对 Amazon EC2 Auto Scaling 文档的重要补充。如需对此文档更新的通知，您可以订阅 RSS 源。

update-history-change	update-history-description	update-history-date
支持新的 Spot 分配策略 (p. 159)	Amazon EC2 Auto Scaling 现在支持新的“容量优化”Spot 分配策略，该策略使用根据可用 Spot 容量选择的最优 Spot 实例池满足您的请求。有关更多信息，请参阅 Amazon EC2 Auto Scaling 用户指南 中的 包括多个实例类型和购买选项的 Auto Scaling 组 。	August 12, 2019
支持标记增强功能 (p. 159)	Amazon EC2 Auto Scaling 现在将标签添加到 Amazon EC2 实例作为启动实例的相同 API 调用的一部分。有关更多信息，请参阅 标记 Auto Scaling 组和实例 。	July 26, 2019
指南更改 (p. 159)	改进了 暂停和恢复扩展流程 主题中的 Amazon EC2 Auto Scaling 文档。更新了 客户托管策略示例 以包括一个示例策略，该策略允许用户仅将特定的后缀自定义服务相关角色传递到 Amazon EC2 Auto Scaling。	June 13, 2019
支持新的 Amazon EBS 功能 (p. 159)	在启动模板主题中增加了对新的 Amazon EBS 功能的支持。在从快照还原时更改 EBS 卷的加密状态。有关更多信息，请参阅 Amazon EC2 Auto Scaling 用户指南 中的 为 Auto Scaling 组创建启动模板 。	May 13, 2019
指南更改 (p. 159)	改进了 Amazon EC2 Auto Scaling 文档中的以下各节： 控制在放大过程中终止哪些 Auto Scaling 实例 、 Auto Scaling 组 、 包括多个实例类型和购买选项的 Auto Scaling 组 以及 Amazon EC2 Auto Scaling 的动态扩展 。	March 12, 2019
支持组合实例类型和购买选项 (p. 159)	在单个 Auto Scaling 组中通过多个购买选项（Spot、按需和预留实例）和实例类型预置和自动扩展实例。有关更多信息，请参阅 Amazon EC2 Auto Scaling 用户指南 中的 包括多个实例类型和购买选项的 Auto Scaling 组 。	November 13, 2018
更新了有关基于 Amazon SQS 扩展的主题 (p. 159)	更新了指南，介绍如何使用自定义指标，根据 Amazon SQS	July 26, 2018

队列中的需求变化来扩展 Auto Scaling 组。有关更多信息，请参阅 Amazon EC2 Auto Scaling 用户指南 中的[基于 Amazon SQS 进行扩展](#)。

下表介绍了在 2018 年 7 月以前对 Amazon EC2 Auto Scaling 文档的一些重要更改。

功能	描述	发行日期
对目标跟踪扩展策略的支持	只需几个步骤就可为应用程序设置动态扩展。有关更多信息，请参阅 Amazon EC2 Auto Scaling 的目标跟踪扩展策略 。	2017 年 7 月 12 日
支持资源级权限	创建 IAM 策略以控制资源级访问。有关更多信息，请参阅 控制对 Amazon EC2 Auto Scaling 资源的访问 。	2017 年 5 月 15 日
监控改进	Auto Scaling 组指标不再需要您启用详细监控。您现在可以启用组指标集合和从控制台中的 Monitoring 选项卡上查看指标图形。有关更多信息，请参阅 使用 Amazon CloudWatch 监控 Auto Scaling 组和实例 。	2016 年 8 月 18 日
支持 应用程序负载均衡器	将一个或多个目标组附加到新的或现有的 Auto Scaling 组。有关更多信息，请参阅 将负载均衡器附加到 Auto Scaling 组 。	2016 年 8 月 11 日
生命周期挂钩的事件	Amazon EC2 Auto Scaling 在执行生命周期挂钩时会向 CloudWatch Events 发送事件。有关更多信息，请参阅 在 Auto Scaling 组扩展时获取 CloudWatch Events 。	2016 年 2 月 24 日
实例保护	阻止 Amazon EC2 Auto Scaling 在缩减时选择终止特定实例。有关更多信息，请参阅 实例保护 。	2015 年 12 月 07 日
分步扩展策略	创建扩展策略，使您能根据警报的严重程度来扩展。有关更多信息，请参阅 扩展策略类型 。	2015 年 7 月 6 日
更新负载均衡器	将负载均衡器附加到现有 Auto Scaling 组或从中分离。有关更多信息，请参阅 将负载均衡器附加到 Auto Scaling 组 。	2015 年 6 月 11 日
支持 ClassicLink	将 Auto Scaling 组中的 EC2-Classic 实例连接到 VPC，让这些链接的 EC2-Classic 实例与 VPC 中的实例能够利用私有 IP 地址进行通信。有关更多信息，请参阅 将 EC2-Classic 实例链接到 VPC 。	2015 年 1 月 19 日
生命周期挂钩	在您对新启动的实例或正在终止的实例执行操作时，请将这些实例保持在等待状态。有关更多信息，请参阅 Amazon EC2 Auto Scaling 生命周期挂钩 。	2014 年 7 月 30 日
分离实例	从 Auto Scaling 组分离实例。有关更多信息，请参阅 从 Auto Scaling 组分离 EC2 实例 。	2014 年 7 月 30 日
将实例置于 Standby 状态	将处于 InService 状态的实例置于 Standby 状态。有关更多信息，请参阅 从 Auto Scaling 组临时删除实例 。	2014 年 7 月 30 日
管理标签	使用 AWS 管理控制台管理您的 Auto Scaling 组。有关更多信息，请参阅 标记 Auto Scaling 组和实例 。	2014 年 5 月 01 日
支持专用实例	通过在创建启动配置时指定部署租期属性启动专用实例。有关更多信息，请参阅 实例部署租期 。	2014 年 4 月 23 日

功能	描述	发行日期
从 EC2 实例创建组或启动配置	使用 EC2 实例创建 Auto Scaling 组或启动配置。有关使用 EC2 实例创建启动配置的信息，请参阅 使用 EC2 实例创建启动配置 。有关使用 EC2 实例创建 Auto Scaling 组的信息，请参阅 使用 EC2 实例创建 Auto Scaling 组 。	2014 年 1 月 02 日
附加实例	您可以通过将 EC2 实例附加到现有 Auto Scaling 组，为该实例启用自动扩展。有关更多信息，请参阅 将 EC2 实例附加到 Auto Scaling 组 。	2014 年 1 月 02 日
查看账户限制	查看对您的账户的 Auto Scaling 资源限制。有关更多信息，请参阅 Auto Scaling 限制 。	2014 年 1 月 02 日
Amazon EC2 Auto Scaling 的控制台支持	使用 AWS 管理控制台访问 Amazon EC2 Auto Scaling。有关更多信息，请参见 Amazon EC2 Auto Scaling 入门 。	2013 年 12 月 10 日
分配公有 IP 地址	将公有 IP 地址分配给启动至 VPC 的实例。有关更多信息，请参阅 在 VPC 中启动 Auto Scaling 实例 。	2013 年 9 月 19 日
实例终止策略	为 Amazon EC2 Auto Scaling 指定实例终止策略，用于终止 EC2 实例。有关更多信息，请参阅 控制在缩减过程中终止哪些 Auto Scaling 实例 。	2012 年 9 月 17 日
对 IAM 角色的支持	用 IAM 实例配置文件启动 EC2 实例。您可以使用此功能为实例指定 IAM 角色，允许您的其他应用程序安全地访问其他 AWS 服务。有关更多信息，请参阅 使用 IAM 角色启动 Auto Scaling 实例 。	2012 年 6 月 11 日
支持 Spot 实例	可以通过在启动配置中指定 Spot 实例出价，在 Auto Scaling 组中请求 Spot 实例。有关更多信息，请参阅 在 Auto Scaling 组中启动 Spot 实例 。	2012 年 6 月 7 日
对组和实例进行标记	标记 Auto Scaling 组并指定将标签也应用于在创建标签之后启动的 EC2 实例。有关更多信息，请参阅 标记 Auto Scaling 组和实例 。	2012 年 1 月 26 日
支持 Amazon SNS	<p>每当 Amazon EC2 Auto Scaling 启动或终止 EC2 实例时，都使用 Amazon SNS 接收通知。有关更多信息，请参阅在 Auto Scaling 组扩展时获取 SNS 通知。</p> <p>Amazon EC2 Auto Scaling 还增加了以下新功能：</p> <ul style="list-style-type: none"> 使用 cron 语法设置重复扩展活动的的能力。有关更多信息，请参阅 PutScheduledUpdateGroupAction API 命令。 增加了新的配置设置，使您无需将已启动的实例添加到负载均衡器 (LoadBalancer) 即可进行扩展。有关更多信息，请参阅 ProcessType API 数据类型。 DeleteAutoScalingGroup 操作中的 ForceDelete 标志指示 Amazon EC2 Auto Scaling 删除 Auto Scaling 组及其关联实例，而无需先等待实例终止。有关更多信息，请参阅 DeleteAutoScalingGroup API 操作。 	2011 年 7 月 20 日
计划扩展操作	增加了对计划扩展操作的支持。有关更多信息，请参阅 计划的 Amazon EC2 Auto Scaling 扩展 。	2010 年 12 月 2 日
支持 Amazon VPC	添加了对 Amazon VPC。有关更多信息，请参阅 在 VPC 中启动 Auto Scaling 实例 。	2010 年 12 月 2 日

功能	描述	发行日期
支持 HPC 集群	增加了对于高性能计算 (HPC) 集群的支持。	2010 年 12 月 2 日
支持运行状况检查	进一步支持对使用 Amazon EC2 Auto Scaling 管理的 EC2 实例进行 Elastic Load Balancing 运行状况检查。有关更多信息，请参阅 结合使用 ELB 运行状况检查和 Auto Scaling 。	2010 年 12 月 2 日
支持 CloudWatch 警报	删除了较旧的触发机制，并重新设计了 Amazon EC2 Auto Scaling 以便使用 CloudWatch 警报功能。有关更多信息，请参阅 适用于 Amazon EC2 Auto Scaling 的动态扩展 。	2010 年 12 月 2 日
暂停和恢复扩展	增加了对暂停和恢复扩展过程的支持。	2010 年 12 月 2 日
支持 IAM	添加了对 IAM。有关更多信息，请参阅 控制对 Amazon EC2 Auto Scaling 资源的访问 。	2010 年 12 月 2 日