AWS Auto Scaling

用户指南

AWS Auto Scaling: 用户指南

Table of Contents

什么是 AWS Auto Scaling？

通过使用 AWS Auto Scaling，您可以在几分钟内为作为您的应用程序的一部分的 AWS 资源配置自动扩展。AWS Auto Scaling 控制台提供了一个单一的用户界面，可以统一管理多个 AWS 服务的自动扩展功能。您可以为单个资源或整个应用程序配置自动扩展。

利用 AWS Auto Scaling，您可以通过扩展计划来配置和管理资源的扩展。扩展计划使用动态扩展和预测式扩展来自动扩展应用程序的资源。这将确保您添加处理应用程序上的负载所需的计算能力然后在不再需要它时进行删除。扩展计划可让您选择扩展策略以定义如何优化资源利用率。您可以针对可用性、成本或这两者的平衡进行优化。此外，您还可以创建自定义扩展策略。

AWS Auto Scaling 对流量存在每日或每周变化的应用程序很有用，这包括：

- 周期性流量，例如正常营业时间内的高资源利用率和夜间的低资源利用率
- 打开和关闭工作负载模式，例如批处理、测试或定期分析
- 可变的流量模式，例如具有峰值增长的营销活动

AWS Auto Scaling 的功能

使用 AWS Auto Scaling 可自动扩展以下资源：

- Amazon EC2 Auto Scaling 组：启动或终止 Auto Scaling 组中的 EC2 实例。
- Amazon EC2 Spot 队列请求：从 Spot 队列请求启动或终止实例，或自动替换由于价格或容量原因而中断的实例。
- Amazon ECS：根据负载变化上调或下调 ECS 服务的预期数量。
- Amazon DynamoDB：启用 DynamoDB 表或全局二级索引以增加或减少其预置的读取和写入容量，从而不受限制地处理流量增加。
- Amazon Aurora：动态调整为 Aurora 数据库集群预配置的 Aurora 只读副本数以处理活动连接或工作负载的变化。

当前可用的扩展功能是动态扩展和预测式扩展。

动态扩展将为您的应用程序中的可扩展资源创建目标跟踪扩展策略。这使您的扩展计划可以根据需要为每个资源增加和删除容量，从而将资源利用率保持在指定的目标值。提供的默认扩展指标基于用于自动扩展的最常用的指标。

预测式扩展的工作方式：

- 负载预测：AWS Auto Scaling 将分析指定负载指标的长达 14 天的历史记录并预测接下来两天的需求。此数据以一小时的间隔提供并且每天都会更新。
- 计划的扩展操作：AWS Auto Scaling 将计划主动增加和删除资源容量的扩展操作以反映负载预测。在计划的时间，AWS Auto Scaling 将使用由计划的扩展操作指定的值更新资源的最小容量。其目的是将资源利用率保持在扩展策略指定的目标值。如果您的应用程序需要的容量大于预期，则可以使用动态扩展来增加额外的容量。
- 最大容量行为：每个资源都有一个最小容量限制和一个最大容量限制，计划的扩展操作指定的值应该在此限制之间。但是，您可以控制在预测容量大于最大容量时应用程序是否可以添加超出其最大容量的资源。

目前，预测式扩展仅适用于 Amazon EC2 Auto Scaling 组。

定价

AWS Auto Scaling 功能由 Amazon CloudWatch 指标和警报启用。如果已支付 CloudWatch 和您使用的其他 AWS 资源的服务费，这些功能将免费提供。

如何开始

有关 AWS Auto Scaling 的简介，我们建议您熟悉以下内容：

- 扩展计划的工作方式 (p. 3)—它介绍了扩展策略、动态扩展和预测式扩展的概念以帮助您熟悉 AWS Auto Scaling。
- AWS Auto Scaling FAQs产品页面上的常见问题提供了有关此服务的好处的信息。—
- 中的 区域和终端节点 AWS General Reference 此表显示 的区域可用性。—AWS Auto Scaling
- Amazon EC2 Auto Scaling 用户指南—本指南向您介绍如何创建和管理要在扩展 Amazon EC2 实例的队组时使用的 Auto Scaling 组。
- Application Auto Scaling 用户指南—本指南为您提供了与超出 Amazon EC2 的容量自动扩展资源相关的主题和资源。当您需要特定于扩展 Amazon EC2 之外的单个可扩展资源或服务的更多信息时，您可以访问本指南中的技术文档。

要开始使用，请完成 开始使用 AWS Auto Scaling (p. 5)中的 AWS Auto Scaling 入门教程。

相关服务

借助 AWS CloudFormation，您可以使用模板（JSON 或 YAML 格式的文本文件）对相关 AWS 资源的集合进行建模和预置。您可以使用 AWS CloudFormation 的示例模板或创建您自己的模板来创建 AWS 资源以及应用程序运行时所需的任何相关依赖项或运行时参数。您还可以使用 AWS CloudFormation 创建扩展计划的模板。

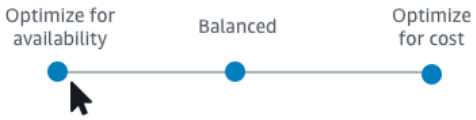
Amazon CloudWatch 是用于 AWS 云资源和您在 AWS 上运行的应用程序的监控服务。CloudWatch 可让您收集和跟踪指标、日志文件和使用警报自动应对应用程序中的变更。您还可使用 AWS CLI 或 API 将自己的自定义指标发送到 CloudWatch。

扩展计划的工作方式

扩展计划是 AWS Auto Scaling 的核心组成部分。在这里，您可以配置一组用于扩展资源的说明。如果您使用 AWS CloudFormation 或向 AWS 资源添加标签，则可以为每个应用程序针对不同的资源集设置扩展计划。AWS Auto Scaling 提供针对每个资源定制的扩展策略的建议。在创建扩展计划后，AWS Auto Scaling 将动态扩展方法和预测式扩展方法相结合以支持您的扩展策略。

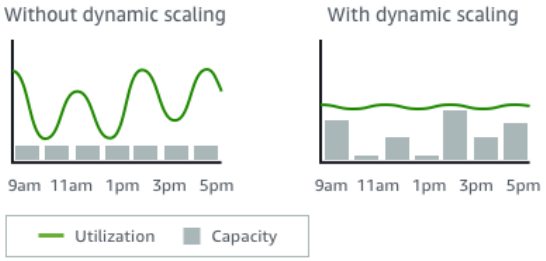
什么是扩展策略？

扩展策略将告知 AWS Auto Scaling 如何在扩展计划中优化资源的利用。您可以针对可用性、成本或这两者的平衡进行优化。或者，您也可以考虑根据自己定义的指标和阈值自行创建自定义策略。您可以为每种资源或资源类型设置单独的策略。



什么是动态扩展？

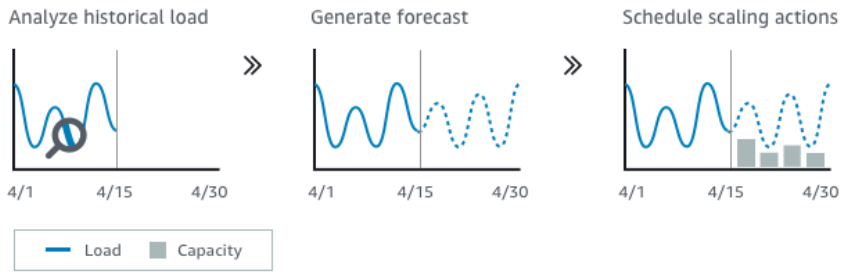
动态扩展为您的扩展计划中的资源创建目标跟踪扩展策略。这些扩展策略将调整资源容量以响应资源利用率的实时变化。其目的是提供足够的容量以将利用率保持在扩展策略指定的目标值。这与恒温器保持家里温度的方式类似。您选择温度，恒温器将完成剩下的工作。



例如，您可以配置扩展计划以使 ECS 服务运行的任务数保持在 CPU 的 75%。当您的服务的 CPU 利用率上升至 75% 以上（这意味着，已使用为服务预留的 CPU 的 75% 以上）时，这将触发您的扩展策略以向您的服务添加另一个任务来帮助处理增加的负载。

什么是预测式扩展？

预测式扩展使用机器学习来分析每个资源的历史负载，并定期预测未来两天的负载。这类似于天气预报的工作方式。利用预测，预测式扩展会生成计划的扩展操作，以确保在应用程序需要之前有资源容量可用。与动态扩展相似，预测式扩展的作用是将利用率保持在扩展策略指定的目标值。



例如，您可以启用预测式扩展并配置您的扩展策略，以将 Auto Scaling 组的平均 CPU 利用率保持在 50%。您的预测要求每天早上 8 点出现流量峰值。您的扩展计划将创建未来的计划扩展操作，以确保您的 Auto Scaling 组已做好提前处理该流量的准备。这有助于使应用程序性能保持不变，目的是始终拥有所需的容量来尽可能让资源利用率保持在接近 50%。

开始使用 AWS Auto Scaling

本节介绍开始使用 AWS Auto Scaling 的步骤。您可以使用 AWS 管理控制台 创建您的第一个扩展计划。然后，了解创建启用了预测式扩展和动态扩展的扩展计划的基础知识。

在创建用于应用程序的扩展计划之前，请全面考察应用程序在 AWS 云中运行时的情况。记录以下内容：

- 您是否具有通过其他控制台创建的现有扩展策略。您可以替换现有的扩展策略，也可以在创建扩展计划时保留这些策略（不允许对策略值进行任何更改）。
- 对您的应用程序中作为整体基于资源的每个可扩展资源有意义的目标利用率。例如，一个 Auto Scaling 组中的 EC2 实例预计使用的 CPU 量与其可用 CPU 相比。或者，对于使用预置吞吐量模型的服务（如 DynamoDB），与可用吞吐量相比，表或索引预期使用的读取和写入活动量。换句话说，消耗的容量与预置容量的比率。创建扩展计划后，您可以随时更改目标利用率。
- 启动和配置服务器需要多长时间。了解这些信息将帮助您为每个 EC2 实例配置一个在启动后预热的窗口，以确保在上个实例仍在启动时不启动新服务器。
- 指标历史是否长到足够用于预测式扩展（如果使用的是新创建的 Auto Scaling 组）。一般而言，具有 14 整天的历史数据将转化为更准确的预测。最小值为 24 小时。

您越了解您的应用程序，您制定扩展计划的效率就越高。

主题

步骤 1：查找您的可扩展资源

在入门部分中，您将创建一个扩展计划并获取通过 AWS 管理控制台使用 AWS Auto Scaling 的实践介绍。

从控制台创建扩展计划时，AWS Auto Scaling 可帮助您查找可扩展资源作为第一步。有三种方法可以从控制台查找新扩展计划的资源：

- 您可以为 AWS CloudFormation 控制台选择 AWS Auto Scaling 堆栈，以便用于自动发现可扩展资源。
- 您可以选择一组标签供 AWS Auto Scaling 控制台用于自动发现可扩展资源。
- 您可以选择要添加到扩展计划的一个或多个 Amazon EC2 Auto Scaling 组。

如果这是您的第一个扩展计划，我们建议您首先选择第三个选项，然后使用 EC2 Auto Scaling 组创建示例扩展计划。

示例扩展计划的先决条件

有关使用控制台创建扩展计划的初学者友好教程，我们建议您首先创建一个 Auto Scaling 组，然后创建扩展计划并添加 Auto Scaling 组。使用 Auto Scaling 组，您可以启用预测式扩展功能和动态扩展功能。您必须启用这两个功能才能使用扩展计划中提供的一整套功能。

如果您还没有 Auto Scaling 组，请首先创建一个。有关更多信息，请参阅 中的 Amazon EC2 Auto Scaling 入门。Amazon EC2 Auto Scaling 用户指南如果您创建了一个新组，则可以随后将其删除。在删除该组后，将停止对它运行的 Amazon EC2 实例收取费用。

按如下所示配置您的 Auto Scaling 组，以确保扩展计划按预期方式工作：

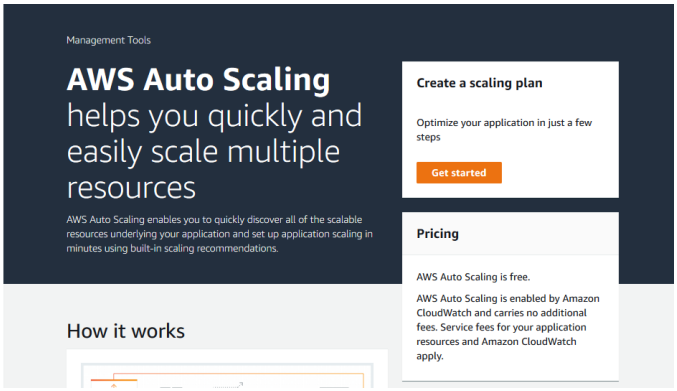
- 在与 Auto Scaling 组关联的启动模板或启动配置中，启用详细监控，从而以 1 分钟的频率获取 EC2 实例的 CloudWatch 指标数据。将收取额外费用。有关更多信息，请参阅 中的Auto Scaling为 实例配置监控。Amazon EC2 Auto Scaling 用户指南
- 启用 Auto Scaling 组指标以获取 CloudWatch 中实例组的聚合数据。有关更多信息，请参阅 中的Auto Scaling启用 组指标。Amazon EC2 Auto Scaling 用户指南
- 如果您使用 T2 或 T3 实例类型，请使用启动模板将实例配置为 `unlimited`，以便它们在您进行测试时能够保持较高的 CPU 性能。可能收取额外费用。有关更多信息，请参阅 https://docs.amazonaws.cn/AWSEC2/latest/UserGuide/burstable-performance-instances-how-to.html#burstable-performance-instances-auto-scaling-grp 中的使用 Auto Scaling 组以“无限”模式启动可突增性能实例Amazon EC2 用户指南（适用于 Linux 实例）。

将 Auto Scaling 组添加到示例扩展计划

现在，您已经创建了 Auto Scaling 组，接下来可以使用 AWS 管理控制台 创建示例扩展计划。

将 Auto Scaling 组添加到新的扩展计划

1. 通过以下网址打开 AWS Auto Scaling 控制台：https://console.amazonaws.cn/awsautoscaling/。
2. 在屏幕顶部的导航栏中，选择在创建 Auto Scaling 组时使用的同一区域。
3. 从欢迎页面中，选择 Get started (开始使用)。

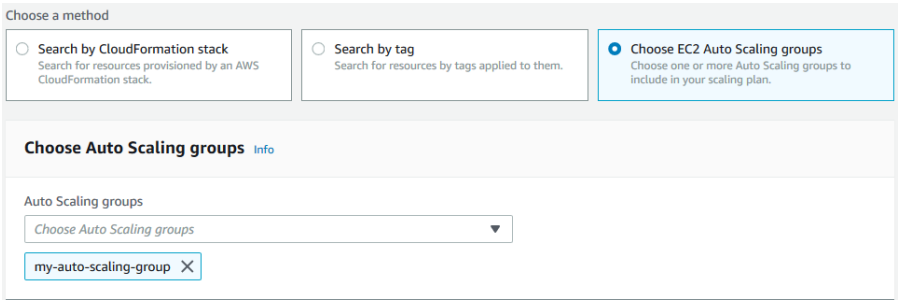


4. 在查找可扩展资源页面上，选择按 CloudFormation 堆栈搜索、按标签搜索或选择 EC2 Auto Scaling 组。

   Note

   本教程假定您选择一个 Auto Scaling 组。稍后，您可以使用此相同过程通过 Search by CloudFormation stack（按 堆栈搜索）或 Search by tag（按标签搜索）选项创建扩展计划。

   - 如果您选择 Search by CloudFormation stack（按 AWS Config 堆栈搜索），请选择要使用的 AWS CloudFormation 堆栈。
   - 如果您选择按标签搜索，则对于每个标签，从键中选择标签键，并从值中选择标签值。要添加标签，请选择 Add another row (添加其他行)。要删除标签，请选择 Remove (删除)。
   - 如果您选择选择 EC2 Auto Scaling 组，然后对于 Auto Scaling 组，选择一个或多个 Auto Scaling 组。



5. 选择 Next （下一步）以将 Auto Scaling 组添加到扩展计划并继续下一步。

   如果您选择了 Search by CloudFormation stack （按 Amazon S3 堆栈搜索）或 Search by tag （按标签搜索）选项，则选择 Next （下一步）将使与堆栈关联的可扩展资源或一组标签可用于扩展计划。当您定义扩展计划时，您接着可以选择要包含或排除其中哪些资源。

了解有关发现可扩展资源的更多信息

如果您已创建示例扩展计划并希望创建更多扩展计划，以下信息将更详细地介绍使用 CloudFormation 堆栈或一组标签的场景。您可以使用此部分决定在使用控制台创建扩展计划时是选择 Search by CloudFormation stack （按 堆栈搜索）还是 Search by tag （按标签搜索）选项来发现可扩展资源。

使用 CloudFormation 堆栈发现可扩展资源

当您使用 CloudFormation 时，使用堆栈来预配置资源。堆栈中的所有资源均由堆栈的模板定义。您的扩展计划在堆栈顶部添加了一个业务流程层，从而可以更轻松地为多个资源配置扩展。如果没有扩展计划，则需要为每个可扩展资源单独设置扩展。这意味着要弄清楚预配置资源和扩展策略的顺序，并了解这些依赖项工作方式的精妙之处。

在 AWS Auto Scaling 控制台中，您可以选择现有堆栈以扫描它是否可配置自动扩展的资源。AWS Auto Scaling 只会查找在所选堆栈中定义的资源。它不会遍历嵌套堆栈。

要确保您的 ECS 服务可在 CloudFormation 堆栈中被发现，AWS Auto Scaling 控制台必须知道哪个 ECS 集群正在运行该服务。这要求您的 ECS 服务与运行该服务的 ECS 集群位于同一 CloudFormation 堆栈中。否则，它们必须是默认集群的一部分。为了正确识别服务，ECS 服务名称在每个 ECS 集群中也必须是唯一的。

有关 CloudFormation 的更多信息，请参阅 中的AWS CloudFormation什么是 ?。AWS CloudFormation 用户指南

使用标签发现可扩展资源

标签提供可用于使用标签筛选条件在 AWS Auto Scaling 控制台中发现相关可扩展资源的元数据。

使用标签查找以下任何资源：

- Aurora 数据库集群
- Auto Scaling组
- DynamoDB 表和全局二级索引

当您按多个标签搜索时，每个资源都必须发现所有列出的标签。

标签可通过许多方式分配。有关更多信息，请参阅 https://docs.amazonaws.cn/general/latest/gr/aws_tagging.html 中的标记 AWS 资源AWS General Reference。

步骤 2：指定扩展策略

使用以下过程为上一步中发现的资源指定扩展策略。

对于每种类型的资源，AWS Auto Scaling 选择最常用于确定在任何给定时间有多少该资源正在使用中的指标。您应选择最合适的扩展策略以根据此指标优化性能。当您启用动态扩展功能和预测式扩展功能时，在它们之间共享扩展策略。有关更多信息，请参阅扩展计划的工作方式 (p. 3)。

有以下扩展策略可用：

- 提高可用性 — AWS Auto Scaling 自动扩展和缩减资源并将资源利用率保持在 40%。当您的应用程序具有紧急且有时无法预测的扩展需求时，此选项很有用。
- 在可用性和成本之间进行平衡 — AWS Auto Scaling 自动扩展和缩减资源并将资源利用率保持在 50%。此选项可帮助您保持高可用性，同时降低成本。
- 成本优化 — AWS Auto Scaling 自动扩展和缩减资源并将资源利用率保持在 70%。如果您的应用程序可以在需求出现意外更改时处理缓冲区容量减少的情况，则此选项可用于降低成本。

例如，扩展计划将您的 Auto Scaling 组配置为根据组中所有实例平均使用的 CPU 量来添加或删除 Amazon EC2 实例。您可选择是否通过更改扩展策略来针对可用性、成本或两者的组合优化使用率。

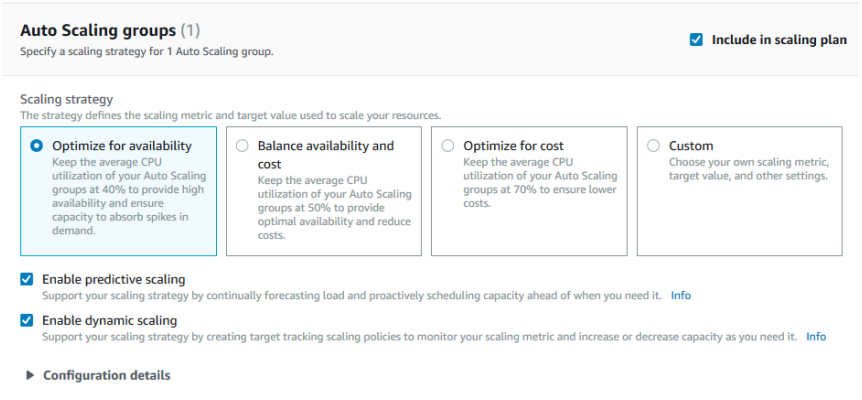
或者，如果现成的策略不能满足您的需求，您可以配置自定义策略。使用自定义策略，您可以更改目标利用率值，选择其他指标，或同时采用这两种方法。

> **Important**
>
> 对于初学者教程来说，完成以下过程的第一步，然后选择下一步继续。（您可以跳过本节的其余部分，因为本教程侧重于使用默认的扩展策略提高可用性，它将 Auto Scaling 组的平均 CPU 利用率保持在 40%。）

指定扩展策略

1. 在 Specify scaling strategy (指定扩展策略) 页上，对于 Scaling plan details (扩展计划详细信息)、Name (名称)，输入扩展计划的名称。扩展计划的名称在您的 AWS 区域扩展计划集中必须是唯一的，最多可包含 128 个字符，并且不得包含竖线“|”、正斜杠“/”或冒号“:”。
2. 对于每种类型的资源，提供以下扩展说明。

   a. 对于 Scaling strategy（扩展策略），请选择以下选项之一：优化可用性、平衡可用性和成本、优化成本或自定义。



   b. 如果您在上一步中选择 Custom (自定义)，请选择 Configuration details (配置详细信息) 下的自定义设置。您可在此处找到可供您使用的指标的列表（如有）和基于 CloudWatch 中数据的相关图表。最近指标历史为图表的重点。

- 对于 Scaling metric (扩展指标)，请选择所需的扩展指标。如果没有其他预定义指标可用，此选项没有可以显示的下拉列表。
- 对于 Target value (目标值)，选择所需的目标利用率值。
- 对于 Load metric (负载指标) [仅限 Auto Scaling 组]，选择适用于预测式扩展的负载指标。
- 对于 Replace external scaling policies (替换外部扩展策略)，请选择是否删除在扩展计划之外（如其他控制台）创建的扩展策略并将其替换为扩展计划创建的新目标跟踪扩展策略。

c. （可选）默认情况下，已为您的 Auto Scaling 组启用预测式扩展。要为您的 Auto Scaling 组禁用预测式扩展，请清除 Enable predictive scaling (启用预测式扩展)。

d. （可选）默认情况下，将为所有资源类型启用动态扩展。要为某种资源禁用动态扩展，请清除启用动态扩展。

e. （可选）默认情况下，当您指定为其发现了多个可扩展资源的应用程序源时，所有资源类型自动包括到您的扩展计划中。要在扩展计划中忽略某种资源，请清除包含在扩展计划中。

3. 完成后，选择 Next。

步骤 3：配置高级设置（可选）

现在您已指定要用于每个资源类型的扩展策略，可以使用配置高级设置步骤，选择按资源自定义任何默认设置。对于每个资源类型，您可以定义多组设置。但是，在大多数情况下，默认设置应该是最佳设置，最小容量和最大容量的值可能例外，这些值应慎重调整。

如果要保留默认设置，则跳过此过程。您可以通过编辑扩展计划随时更改这些设置。

> Important
>
> 在初学者教程中，我们可以进行一些更改，更新您 Auto Scaling 组的最大容量并启用仅预测模式的预测式扩展。虽然您不需要自定义教程的所有设置，我们可以简单查看一下各个部分中的设置。

常规设置

使用此过程可以按照各个资源，查看和自定义您在上一步中指定的设置。您还可以自定义每个资源的最小容量和最大容量。

查看和自定义常规设置

1. 在配置高级设置页面上，选择左侧任意部分标题的箭头以展开该部分。在本教程中，展开Auto Scaling 组部分。
2. 从显示的表中，选择您在本教程中使用的 Auto Scaling 组。
3. 保留选中包含在扩展计划中选项。如果未选择此选项，则扩展计划中会忽略资源。如果您未包含至少一个资源，则无法创建扩展计划。
4. 要展开视图并查看常规设置部分的详细信息，请选择部分标题左侧的箭头。
5. 您可以选择以下任意项。在本教程中，找到最大容量设置并输入值 3 代替当前值。

- 扩展策略 — 允许您提高可用性、优化成本，或使可用性和成本达到平衡，或指定自定义策略。
- 启用动态扩展 — 如果清除了此设置，则无法使用目标跟踪扩展配置扩展所选资源。
- 启用预测式扩展 — [仅限 Auto Scaling 组] 如果清除此设置，则无法使用预测式扩展来扩展所选组。
- Scaling metric (扩展指标) — 指定要使用的扩展指标。如果您选择 Custom（自定义），则可以指定要使用的自定义指标，而不是控制台中可用的预定义指标。有关更多信息，请参阅此部分中的下一个主题。
- Target value (目标值) — 指定要使用的目标利用率值。
- Load metric (负载指标) — [仅限 Auto Scaling 组] 指定要使用的负载指标。如果您选择 Custom（自定义），则可以指定要使用的自定义指标，而不是控制台中可用的预定义指标。有关更多信息，请参阅此部分中的下一个主题。

- 最小容量 — 指定资源的最小容量。AWS Auto Scaling 可确保您的资源永远不会低于这个数量。
- 最大容量 — 指定资源的最大容量。AWS Auto Scaling 可确保您的资源永远不会高于这个数量。

  Note

  使用预测式扩展时，您也可以选择根据预测容量来使用其他最大容量行为。此设置位于预测式扩展设置部分中。

自定义指标

AWS Auto Scaling 提供最常用于自动扩展的指标。但是，根据您的需求，您可能偏爱从不同的指标而不是控制台中的指标获取数据。Amazon CloudWatch 具有许多不同的指标可供选择。CloudWatch 还允许您发布自己的指标。

您可以使用 JSON 指定 CloudWatch 自定义指标。在按照这些说明进行操作之前，建议您熟悉一下 Amazon CloudWatch 用户指南。

要指定自定义指标，您可以使用模板中的一组必需参数构造 JSON 格式的负载。您为 CloudWatch 中的每个参数添加值。在扩展计划的高级设置中，我们提供模板作为扩展指标和负载指标的自定义选项的一部分。

JSON 通过两种方式表示数据：

- 对象，其是无序名称-值对集合。对象是在左大括号 ({) 和右大括号 (}) 内定义的。每个名称-值对以名称开头，后接一个冒号，再接值。名称-值对是用逗号隔开的。
- 数组，其是有序值集合。数组是在左方括号 ([) 和右方括号 (]) 内定义的。数组中的项目是用逗号隔开的。

下面是为每个参数提供示例值的 JSON 模板的示例：

```
{
  "MetricName": "MyBackendCPU",
  "Namespace": "MyNamespace",
  "Dimensions": [
    {
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }
  ],
  "Statistic": "Sum"
}
```

有关更多信息，请参阅 中的自定义扩展指标规范和自定义负载指标规范。AWS Auto Scaling API 参考

动态扩展设置

使用此过程可以查看和自定义 AWS Auto Scaling 创建的目标跟踪扩展策略的设置。

查看和自定义动态扩展的设置

1. 要展开视图并查看动态扩展设置部分的详细信息，请选择部分标题左侧的箭头。
2. 您可以为以下项进行选择。但是，默认设置非常适用于本教程。

   - 替换外部扩展策略 — 如果清除此设置，则将保留在扩展计划外创建的现有扩展策略并且不会创建新的扩展策略。
   - 禁用缩减 — 如果清除此设置，则在指定指标低于目标值时，允许自动缩减以减小资源的当前容量。
   - Cooldown (冷却) — 创建扩大和缩减冷却时间。冷却时间是扩展策略等待上一扩展活动生效的时间量。有关更多信息，请参阅 https://docs.amazonaws.cn/autoscaling/application/userguide/

application-auto-scaling-target-tracking.html#target-tracking-cooldown 中的冷却时间Application Auto Scaling 用户指南。（如果资源是 Auto Scaling 组，则不会显示此设置。）

- 实例预热 — [仅限 Auto Scaling 组] 控制新启动的实例在多长时间后开始作用于 CloudWatch 指标。有关更多信息，请参阅 https://docs.amazonaws.cn/autoscaling/ec2/userguide/as-scaling-target-tracking.html#as-target-tracking-scaling-warmup 中的实例预热Amazon EC2 Auto Scaling 用户指南。

预测性扩展设置

如果您的资源是 Auto Scaling 组，请使用此过程来查看并自定义 AWS Auto Scaling 用于预测式扩展的设置。

查看和自定义预测式扩展的设置

1. 要展开视图并查看预测式扩展设置部分的详细信息，请选择部分标题左侧的箭头。
2. 您可以为以下项进行选择。在本教程中，请将预测式扩展模式更改为仅预测。
   - Predictive scaling mode (预测式扩展模式) — 指定扩展模式。默认值为 Forecast and scale (预测和扩展)。如果您将它更改为仅预测，则扩展计划将预测未来容量，但不会应用扩展操作。
   - 预启动实例 — 调整扩大时要提前运行的扩展操作。例如，预测表示在上午 10:00 点增加容量，缓冲时间为 5 分钟（300 秒）。这样，对应的扩展操作的运行时间为上午 9:55。这对于 Auto Scaling 组很有帮助，这些组在从实例启动到服务可能需要几分钟。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。默认值为 300 秒。
   - 最大容量行为 — 控制当预测容量接近或超过当前指定的最大容量时，所选资源是否可以扩展到最大容量以上。默认值为强制实施最大容量设置。
     - 强制实施最大容量设置 — AWS Auto Scaling 无法将资源容量扩展到高于最大容量。最大容量是作为硬限制实施的。
     - 将最大容量设置为等于预测容量—AWS Auto Scaling 可以将资源容量扩展到高于最大容量，直至等于但不能超过预测容量。
     - 提高最大容量以超过预测容量 — AWS Auto Scaling 可以按指定的缓冲区值扩展资源容量来超过最大容量。目的是在出现意外流量时，为目标跟踪扩展策略提供额外的容量。
   - 最大容量行为缓冲区 — 如果您选择提高最大容量以超过预测容量，选择在预测容量接近或超过最大容量时，所用容量缓冲区的大小。该值是作为相对于预测容量的百分比指定的。例如,使用 10% 的缓冲区，如果预测容量为 50，最大容量为 40，则有效的最大容量是 55。
3. 自定义完设置之后，选择 Next (下一步)。

   Note

   要还原您的任何更改，请选择所需资源，然后选择 Revert to original (还原为最初设置)。这会将所选资源重置为扩展计划中的上一个已知状态。

步骤 4：创建您的扩展计划

在 Review and create (审核和创建) 页面上，审核您的扩展计划并选择 Create scaling plan (创建扩展计划)。您会定向到显示扩展计划状态的页面。在更新资源时，扩展计划的创建可能需要一点时间才能完成。

使用预测式扩展时，AWS Auto Scaling 将分析过去 14 天的指定负载指标的历史记录（至少需要 24 小时的数据）以生成对未来两天的预测。然后，它将安排扩展操作来调整资源容量调整，使之与预测期内每小时的预测匹配。

在扩展计划创建完成之后，通过在扩展计划屏幕中选择其名称来查看扩展计划详细信息。

（可选）查看资源的扩展信息

使用此过程可以查看为资源创建的扩展信息。

数据通过以下方式提供：

- 显示 CloudWatch 中的最近指标历史数据的图表。
- 预测式扩展图显示根据 AWS Auto Scaling. 的数据进行的负载预测以及容量预测。
- 表中列出了为资源计划的所有预测式扩展操作。

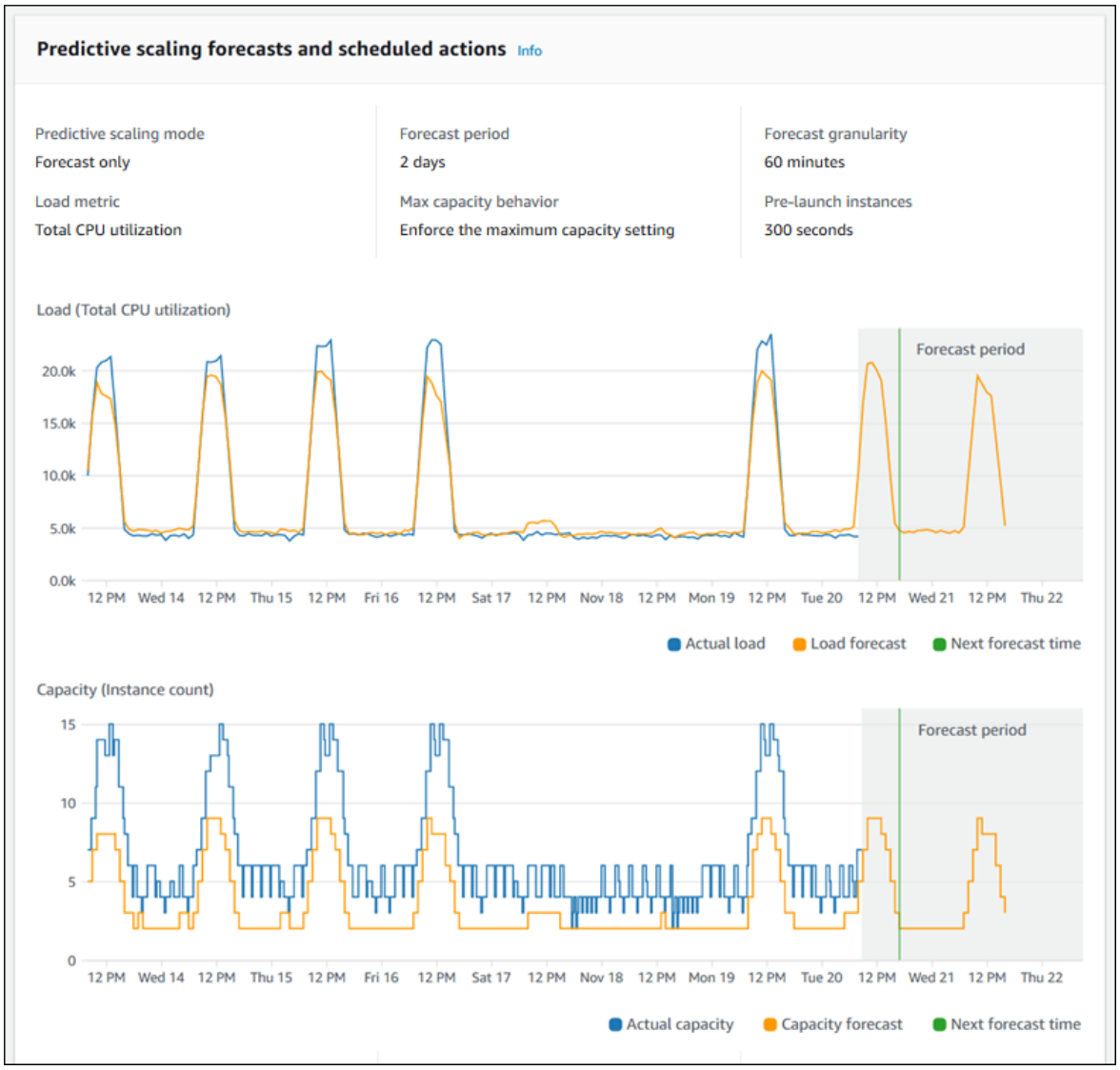
查看资源的扩展信息

1. 通过以下网址打开 AWS Auto Scaling 控制台：https://console.amazonaws.cn/awsautoscaling/。
2. 在 Scaling plans (扩展计划) 页面上，选择扩展计划。
3. 在 Scaling plan details (扩展计划详细信息) 页面上，选择要查看的资源。

监控和评估预测

当扩展计划启动运行时，您可以监控负载预测、容量预测和扩展操作，以检查预测式扩展的性能。所有这类数据均可在 AWS Auto Scaling 控制台中提供，适用于所有启用了预测式扩展的 Auto Scaling 组。请记住，您的扩展计划需要至少 24 小时的历史负载数据来进行初次预测。

在以下示例中，每个图表的左侧都显示历史模式。右侧显示扩展计划在预测期间生成的预测。实际值和预测值（分别为蓝色和橙色）均绘制。



AWS Auto Scaling 自动从您的数据中学习。首先，它会进行负载预测。然后，容量预测计算确定支持应用程序所需的最小实例数。根据容量预测，AWS Auto Scaling 计划在预测的负载变化之前扩展 Auto Scaling 组的扩展操作。如果启用了动态扩展（推荐），则 Auto Scaling 组可以根据实例组的当前利用率扩展其他容量（或删除容量）。

当评估预测式扩展的执行情况时，可监控实际值和预测值在一段时间内 的接近程度。当您创建扩展计划时，AWS Auto Scaling 将提供基于最新实际值的图表。它还提供接下来 48 小时内的初始预测。但是，在创建扩展计划后，几乎没有可与实际数据进行比较的预测数据。请等到扩展计划已获取若干时间段的预测值，然后再将历史预测值与实际值进行比较。经过几天的每日预测后，您将有更多的预测值样本与实际值进行比较。

对于每天发生的模式，创建扩展计划和评估预测有效性之间的时间间隔可以短至为几天。但是，此时间长度不足以基于最近模式更改来评估预测。例如，假设您正在查看对某个 Auto Scaling 组的预测，该组在过去一周启动了一个新的市场营销活动。该活动显著增加了您在每周的相同两天的 Web 流量。在类似这样的情况下，我们建议您等待该组收集完整的一周或两周的新数据，然后再评估预测的有效性。对于仅仅开始收集指标数据的全新 Auto Scaling 组，上述建议同样适用。

如果您在监控实际值和预测值一段时间之后，发现它们并不匹配，则还应考虑负载指标的选择。若要有效发挥作用，负载指标必须表示对 Auto Scaling 组中所有实例的总负载的可靠而准确的度量。负载指标是预测性

扩展的核心。如果您选择非最佳负载指标，则它可能会阻止预测性扩展，从而进行准确的负载和容量预测，并为您的 Auto Scaling 组安排正确的容量调整。

步骤 5：清除

完成入门教程后，您可以选择保留您的扩展计划。但是，如果您的扩展计划未在活跃使用中，则应考虑将其删除以免您的账户产生不必要的费用。

删除扩展计划将删除目标跟踪扩展策略、其关联 CloudWatch 警报和 AWS Auto Scaling 代表您创建的预测式扩展操作。

删除扩展计划不会删除您的 AWS CloudFormation 堆栈、Auto Scaling 组或其他可扩展资源。

删除扩展计划

1. 通过以下网址打开 AWS Auto Scaling 控制台：https://console.amazonaws.cn/awsautoscaling/。
2. 在扩展计划页面上，选择您为此教程创建的扩展计划，然后选择删除。
3. 当系统提示进行确认时，选择 Delete。

在您删除扩展计划后，您的资源不会恢复到其原始容量。例如，如果您的 Auto Scaling 组在您删除扩展计划时扩展到 10 个实例，则您的组在扩展计划删除后仍将扩展到 10 个实例。您可以通过分别访问各个服务的控制台，更新特定资源的容量。

删除 Auto Scaling 组

为了防止您的账户产生 Amazon EC2 费用，您还应删除为本教程创建的 Auto Scaling 组。

有关分步说明，请参阅 Auto Scaling 中的删除 组Amazon EC2 Auto Scaling 用户指南。

步骤 6：后续步骤

现在，您已经熟悉了 AWS Auto Scaling 及其部分功能，您可能希望尝试使用 AWS CloudFormation 创建自己的扩展计划模板。

AWS CloudFormation 模板是 JSON 或 YAML 格式的文本文件，描述了运行应用程序或服务所需的 AWS 基础设施以及基础设施组件之间的任何互连。借助 AWS CloudFormation，您可以将关联的资源集合作为 堆栈 进行部署和管理。AWS CloudFormation 不收取额外费用，您只需为运行应用程序所需的 AWS 资源付费。资源可包括您在模板中定义的任何 AWS 资源。有关更多信息，请参阅 AWS CloudFormation 中的 概念。AWS CloudFormation 用户指南

在 AWS CloudFormation 用户指南 中，我们提供了一个简单的模板帮助您入门。示例模板在 模板参考文档的 AWS：：AutoScalingPlans：：ScalingPlanAWS CloudFormation 部分中作为一个示例提供。示例模板为单个 Auto Scaling 组创建扩展计划，并启用预测式扩展和动态扩展。

有关更多信息，请参阅 中的 AWS CloudFormation 入门。AWS CloudFormation 用户指南

扩展计划的最佳实践AWS Auto Scaling

以下最佳实践可帮助您充分利用扩展计划：

- 只要可能，您应按照具有 1 分钟频率的 Amazon EC2 实例指标扩展，因为这可以确保更快地响应利用率变化。按具有 5 分钟频率的指标扩展会导致响应时间变慢，并且可能导致按过期的指标数据扩展。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。对于其他费用，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 中的Auto Scaling为实例配置监控。Amazon EC2 Auto Scaling 用户指南
- 此外，建议您启用 Auto Scaling 组指标。否则，实际容量数据不会显示在完成“创建扩展计划”向导后提供的容量预测图中。要启用 Auto Scaling 组指标，请在 Amazon EC2 控制台中打开 Auto Scaling 组，并从 Monitoring (监控) 选项卡中，选择 Enable Group Metrics Collection (启用组指标集合)。这些指标对组加以描述，而非其任何实例。有关更多信息，请参阅 中的Auto Scaling启用 组指标。Amazon EC2 Auto Scaling 用户指南
- 检查您的 Auto Scaling 组使用的实例类型。具有可突增性能的 Amazon EC2 实例（T3 和 T2 实例）设计为提供基准级别的 CPU 性能，并且可在您的工作负载需要时突增至更高的级别。根据扩展计划指定的目标利用率，您可以管理超出基准的风险，然后用完 CPU 积分，这将限制性能。有关更多信息，请参阅可突增性能实例的 CPU 积分和基准性能。要将这些实例配置为 `unlimited`，请参阅 中的使用 Auto Scaling 组以“无限”模式启动可突增性能实例。Amazon EC2 用户指南（适用于 Linux 实例）

其他考虑因素

考虑以下其他注意事项：

- 预测式扩展使用工作负载预测来计划未来的容量。预测质量因工作负载的周期性和所训练预测模型的适用性而异。可以在仅预测模式下运行预测式扩展，以评估预测的质量和预测创建的扩展操作。您可以在创建扩展计划时将预测式扩展模式设置为仅预测，然后在完成评估预测质量后将其更改为预测和缩放。有关更多信息，请参阅 预测性扩展设置 (p. 11) 和 监控和评估预测 (p. 12)。
- 如果您选择为预测式扩展指定不同的指标，则必须确保扩展指标和负载指标密切相关。指标值必须随着 Auto Scaling 组中实例的数量按比例增加和缩小。这样可确保指标数据可用于随实例数量按比例扩展或缩减。例如，负载指标是请求计数总计，扩展指标是平均 CPU 利用率。如果请求计数总计增加了 50%，则还应将平均 CPU 利用率增加 50%，前提是容量保持不变。
- 在创建您的扩展计划之前，您应删除之前的任何计划扩展操作（前提是，您访问创建扩展策略的控制台不再需要这些操作）。AWS Auto Scaling 不会创建与现有计划扩展操作重叠的预测式扩展操作。
- 您的最小容量和最大容量的自定义设置，以及用于动态扩展的其他设置将显示在其他控制台中。但是，我们建议，您在创建扩展计划后，不要通过其他控制台修改这些设置，因为您的扩展计划不从其他控制台接收更新。
- 您的扩展计划可以包含来自多个服务的资源，但每个资源一次只能在一个扩展计划中。

避免 ActiveWithProblems 错误

当创建扩展计划或将资源添加到扩展计划时，可能会出现“ActiveWithProblems”错误。当扩展计划处于活动状态，但一个或多个资源的扩展配置无法应用时，出现此错误。

通常情况下，这是因为资源已经具有扩展策略，或 Auto Scaling 组不符合预测式扩展的最低要求。

如果您的任何资源已经具有来自各种 AWS 控制台的扩展策略，则默认情况下，AWS Auto Scaling 不会覆盖这些其他扩展策略或创建新的扩展策略。您可以选择删除现有扩展策略，并将其替换为从 AWS Auto Scaling 控制台创建的目标跟踪扩展策略。为此，您可以为每个具有要覆盖的扩展策略的资源启用 Replace external scaling policies (替换外部扩展策略) 设置。

对于预测性扩展，在创建新的 Auto Scaling 组来配置预测式扩展后，我们建议等待 24 小时时间。至少必须有 24 小时的历史数据才能生成初始预测。如果该组具有少于 24 小时的历史数据并且启用了预测式扩展，则会导致扩展计划在该组收集所需数据量之后的下一个预测期之前无法生成预测数据。但是，您也可以编辑和保存扩展计划，以便在 24 小时的数据可用后立即重新启动预测过程。

中的安全性AWS Auto Scaling

云安全性 AWS 是最高优先级。作为 AWS 客户，您将从专为满足大多数安全敏感型组织的要求而打造的数据中心和网络架构中受益。

安全性是 AWS 和您的共同责任。TheThethe 共同责任模式 将这种情况描述为安全性 的 云和安全 在 云:

- 云安全 – AWS 负责保护运行的基础设施 AWS 服务 AWS 云。 AWS 还为您提供可安全使用的服务。第三方审计师定期测试并验证我们的安全有效性 AWS 合规计划. 了解适用于 AWS Auto Scaling，参见 合规计划中的AWS服务.
- 云中的安全性 – 您的责任由 AWS 您使用的服务。您还需要对其他因素负责，包括您的数据的敏感性、您公司的要求以及适用的法律法规。

该文档帮助您了解如何在使用 时应用责任共担模型。AWS Auto Scaling. 以下主题说明如何配置 AWS Auto Scaling 以实现您的安全性和合规性目标。您还将了解如何使用其他 AWS 服务来帮助您监控和保护 AWS Auto Scaling 资源。

主题

AWS Auto Scaling 和数据保护

AWS 责任共担模式适用于 AWS Auto Scaling 中的数据保护。如该模式中所述，AWS 负责保护运行所有 AWS 云的全球基础设施。您负责维护对托管在此基础设施上的内容的控制。此内容包括您所使用的 AWS 服务的安全配置和管理任务。有关数据隐私的更多信息，请参阅数据隐私常见问题。

出于数据保护目的，我们建议您保护 AWS 账户凭证并使用 AWS Identity and Access Management (IAM) 设置单独的用户账户。这仅向每个用户授予履行其工作职责所需的权限。我们还建议您通过以下方式保护您的数据：

- 对每个账户使用 Multi-Factor Authentication (MFA)。
- 使用 SSL/TLS 与 AWS 资源进行通信。建议使用 TLS 1.2 或更高版本。
- 使用 AWS CloudTrail 设置 API 和用户活动日志记录。
- 使用 AWS 加密解决方案以及 AWS 服务中的所有默认安全控制。
- 使用高级托管安全服务（例如 Amazon Macie），它有助于发现和保护存储在 Amazon S3 中的个人数据。
- 如果在通过命令行界面或 API 访问 AWS 时需要经过 FIPS 140-2 验证的加密模块，请使用 FIPS 终端节点。有关可用的 FIPS 终端节点的更多信息，请参阅美国联邦信息处理标准 (FIPS) 第 140-2 版。

我们强烈建议您切勿将敏感的可识别信息（例如您客户的账号）放入自由格式字段（例如 Name (名称) 字段）。这包括使用控制台、API、AWS CLI 或 AWS 开发工具包处理 AWS Auto Scaling 或其他 AWS 服务

时。您输入到 AWS Auto Scaling 或其他服务中的任何数据都可能被选取以包含在诊断日志中。当您向外部服务器提供 URL 时，请勿在 URL 中包含凭证信息来验证您对该服务器的请求。

适用于 AWS Auto Scaling 的 Identity and Access Management

AWS Identity and Access Management (IAM) 是一项 AWS 服务，可帮助管理员安全地控制对 AWS 资源的访问。IAM 管理员控制谁可以通过身份验证（登录）和授权（具有权限）以使用 AWS Auto Scaling 资源。IAM 是一项无需额外费用即可使用的 AWS 服务。

要使用 AWS Auto Scaling，您需要 AWS 账户和 AWS 凭证。为了提高 AWS 账户的安全性，我们建议您使用 IAM 用户（而不使用 AWS 账户凭证）来提供访问凭证。有关更多信息，请参阅 AWS 一般参考 中的 AWS 账户根用户凭证与 IAM 用户凭证和 IAM 用户指南 中的 IAM 最佳实践。

有关 IAM 用户以及它们对于账户的安全性为何十分重要的概述，请参阅 AWS 一般参考 中的 AWS 安全凭证。

有关使用 IAM 的详细信息，请参阅 IAM 用户指南。

访问控制

您可以使用有效的凭证来对自己的请求进行身份验证，但您还必须拥有权限才能创建或访问 AWS Auto Scaling 资源。例如，您必须具有创建扩展计划、配置预测式扩展等的权限。

以下部分提供了详细信息来说明 IAM 管理员如何使用 IAM 控制哪些用户可执行 AWS Auto Scaling 操作，从而对您的 AWS 资源进行保护。

主题

AWS Auto Scaling 如何与 IAM 协同工作

在使用 IAM 管理对 AWS Auto Scaling 的访问之前，您应了解哪些 IAM 功能可与 AWS Auto Scaling 协同工作。要概括了解 AWS Auto Scaling 及其他 AWS 服务如何与 IAM 协同工作，请参阅 AWS 中的IAM可与 协同工作的 IAM 用户指南 服务。

主题

AWS Auto Scaling 基于身份的策略

使用 IAM 基于身份的策略，您可以指定允许或拒绝的操作和资源，并指定在什么条件下允许或拒绝操作。AWS Auto Scaling 支持特定操作、资源和条件键。要了解您在 JSON 策略中使用的所有元素，请参阅 IAM 中的 JSON 策略元素参考。IAM 用户指南

Actions

Administrators can use AWS JSON policies to specify who has access to what. That is, which principal can perform actions on what resources, and under what conditions.

JSON 策略的 `Action` 元素描述可用于在策略中允许或拒绝访问的操作。策略操作通常与关联的 AWS API 操作同名。有一些例外情况，例如没有匹配 API 操作的仅限权限 操作。还有一些操作需要在策略中执行多个操作。这些附加操作称为相关操作。

在策略中包含操作以授予执行相关操作的权限。

中的策略操作在操作前使用以下前缀：AWS Auto Scaling`autoscaling-plans:`。 策略语句必须包括 `Action` 或 `NotAction` 元素。AWS Auto Scaling 定义了自己的一组操作，这些操作描述了您可以使用该服务执行的任务。

要在单个语句中指定多项操作，请使用逗号将它们隔开，如下例所示。

```
"Action": [
      "autoscaling-plans:DescribeScalingPlans",
      "autoscaling-plans:DescribeScalingPlanResources"
```

您也可以使用通配符 (*) 指定多个操作。例如，要指定以单词 `Describe` 开头的所有操作，请包括以下操作。

```
"Action": "autoscaling-plans:Describe*"
```

要查看 AWS Auto Scaling 操作的列表，请参阅 https://docs.amazonaws.cn/autoscaling/plans/APIReference/API_Operations.html 中的操作AWS Auto Scaling API 参考。

Resources

`Resource` 元素指定要向其应用操作的对象。

AWS Auto Scaling 没有可用作 `Resource` 策略声明的 IAM 元素的服务定义的资源。因此，IAM 策略中没有适用于 AWS Auto Scaling 的 Amazon 资源名称 (ARN)。要控制对 AWS Auto Scaling 操作的访问，请在编写 IAM 策略时始终使用 * (星号) 作为资源。

条件键

在 `Condition` 元素或 `Condition` 块中，可以指定语句生效的条件。例如，您可能希望策略仅在特定日期后应用。要表示条件，请使用预定义的条件键。

AWS Auto Scaling 不提供任何特定于服务的条件键，但支持使用某些全局条件键。要查看所有 AWS 全局条件键，请参阅 中的 AWS 全局条件上下文键。IAM 用户指南

`Condition` 元素是可选的。

Examples

要查看 AWS Auto Scaling 基于身份的策略的示例，请参阅AWS Auto Scaling 基于身份的策略示例 (p. 22)。

AWS Auto Scaling 基于资源的策略

其他 AWS 服务 (如 Amazon Simple Storage Service) 也支持基于资源的权限策略。例如，您可以将权限策略挂载到 S3 存储桶以管理对该存储桶的访问权限。

AWS Auto Scaling 不支持基于资源的策略。

访问控制列表 (ACL)

AWS Auto Scaling 不支持访问控制列表 (ACL)。

基于 AWS Auto Scaling 标签的授权

AWS Auto Scaling 没有可以标记的服务定义的资源。因此，它不支持基于标签控制访问。

AWS Auto Scaling IAM 角色

IAM 角色是 AWS 账户中具有特定权限的实体。

将临时凭证用于 AWS Auto Scaling

您可以使用临时凭证进行联合身份登录、代入 IAM 角色或代入跨账户角色。您可以通过调用 AWS STS API 操作（如 AssumeRole 或 GetFederationToken）来获取临时安全凭证。

AWS Auto Scaling 支持使用临时凭证。

服务相关角色

AWS Auto Scaling 支持服务相关角色。

根据您向扩展计划添加的资源，此功能会自动为 Amazon EC2 Auto Scaling 创建一个服务相关角色，并为 Application Auto Scaling 的每个资源类型创建一个服务相关角色。

如果您启用预测式扩展，此功能会自动创建 AWS Auto Scaling 服务相关角色。

有关更多信息，请参阅适用的角色特定文档：

- 中的 Amazon EC2 Auto Scaling 的服务相关角色Amazon EC2 Auto Scaling 用户指南
- 中的 Application Auto Scaling 的服务相关角色Application Auto Scaling 用户指南
- 本指南中的AWS Auto Scaling 服务相关角色 (p. 20)

您可以使用以下过程来确定您的账户是否已经具有服务相关角色。

确定服务相关角色是否已存在

1. 通过以下网址打开 IAM 控制台：https://console.amazonaws.cn/iam/。
2. 在导航窗格中，选择 Roles。
3. 在列表中搜索“AWSServiceRole”以查找您账户中存在的服务相关角色。查找您要检查的服务相关角色的名称。

服务角色

AWS Auto Scaling 没有服务角色。

AWS Auto Scaling 服务相关角色

> **Important**
>
> 要获取有关使用扩展计划所需的服务相关角色的完整信息，请参阅服务相关角色 (p. 20)。

AWS 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。服务相关的角色是一种与 AWS 服务直接关联的独特类型的 AWS Identity and Access Management (IAM) 角色。

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 https://docs.amazonaws.cn/IAM/latest/UserGuide/using-service-linked-roles.html 中的使用服务相关角色IAM 用户指南。

下面的部分介绍如何创建和管理 AWS Auto Scaling 服务相关角色。首先配置权限以允许 IAM 实体（如用户、组或角色）创建、编辑或删除服务相关角色。

服务相关角色授予的权限

AWS Auto Scaling 使用 _EC2AutoScalingAWSServiceRoleForAutoScalingPlans 服务相关角色来代表您管理 组的预测式扩展。Amazon EC2 Auto Scaling此角色预定义了代表您进行以下调用的权限：

- `cloudwatch:GetMetricData`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:DescribeScheduledActions`
- `autoscaling:BatchPutScheduledUpdateGroupAction`
- `autoscaling:BatchDeleteScheduledAction`

此角色信任 `autoscaling-plans.amazonaws.com` 服务来代入它。

创建服务相关角色（自动）

AWS Auto Scaling 在首次创建启用了预测式扩展的扩展计划时为您创建 _EC2AutoScalingAWSServiceRoleForAutoScalingPlans 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 https://docs.amazonaws.cn/IAM/latest/UserGuide/using-service-linked-roles.html#service-linked-role-permissions 中的服务相关角色权限IAM 用户指南或本指南中有关 的信息。创建服务相关角色所需的权限 (p. 25)

创建服务相关角色（手动）

要手动创建服务相关角色，您可以使用 IAM 控制台、AWS CLI 或 IAM API。有关更多信息，请参阅 https://docs.amazonaws.cn/IAM/latest/UserGuide/using-service-linked-roles.html#create-service-linked-role 中的创建服务相关角色IAM 用户指南。

创建服务相关角色 (AWS CLI)

使用以下 create-service-linked-role CLI 命令创建 AWS Auto Scaling 服务相关角色。

```
aws iam create-service-linked-role --aws-service-name autoscaling-plans.amazonaws.com
```

编辑服务相关角色

利用 创建的 AWSServiceRoleForAutoScalingPlans_EC2AutoScaling 角色，您可以仅编辑其描述而不是其权限。AWS Auto Scaling有关更多信息，请参阅 https://docs.amazonaws.cn/IAM/latest/UserGuide/using-service-linked-roles.html#edit-service-linked-role 中的编辑服务相关角色IAM 用户指南。

删除服务相关角色

如果您不再使用扩展计划，我们建议您删除服务相关角色。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扩展计划结合使用，您必须先删除启用了预测式扩展的所有扩展计划，然后才能删除该角色。这可防止您无意中删除您管理扩展计划所需的权限，从而保护您的扩展计划。有关更多信息，请参阅步骤 5：清除 (p. 14)。

您可以使用 IAM 删除服务相关角色。有关更多信息，请参阅 https://docs.amazonaws.cn/IAM/latest/UserGuide/using-service-linked-roles.html#delete-service-linked-role 中的删除服务相关角色IAM 用户指南。

在删除 AWSServiceRoleForAutoScalingPlans_EC2AutoScaling 服务相关角色后，当您创建启用了预测式扩展的扩展计划时，AWS Auto Scaling 将再次创建该角色。

服务相关角色的受支持区域

AWS 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。

AWS Auto Scaling 基于身份的策略示例

默认情况下，全新的 IAM 用户没有执行任何操作的权限。IAM 管理员必须创建 IAM 策略，以便为用户和角色授予执行 AWS Auto Scaling 操作（如配置扩展策略）的权限。然后，管理员必须将这些策略附加到需要这些权限的 IAM 用户或角色。

要了解如何使用这些示例 JSON 策略文档创建 IAM 策略，请参阅 中的在 JSON 选项卡上创建策略。IAM 用户指南

如果您是首次创建策略，建议您先在账户中创建 IAM 用户并按顺序将策略附加到用户。在将每个策略附加到用户时，可使用控制台验证该策略的效果。

主题

策略最佳实践

基于身份的策略非常强大。它们确定某个人是否可以创建、访问或删除您账户中的 AWS Auto Scaling 资源。这些操作可能会使 AWS 账户产生成本。创建或编辑基于身份的策略时，请遵循以下准则和建议：

- 开始使用 AWS 托管策略 – 要快速开始使用 AWS Auto Scaling，请使用 AWS 托管策略，为您的员工授予他们所需的权限。这些策略已在您的账户中提供，并由 AWS 维护和更新。有关更多信息，请参阅 IAM 用户指南 中的利用 AWS 托管策略开始使用权限。
- 授予最低权限 – 创建自定义策略时，仅授予执行任务所需的许可。最开始只授予最低权限，然后根据需要授予其他权限。这样做比起一开始就授予过于宽松的权限而后再尝试收紧权限来说更为安全。有关更多信息，请参阅 IAM 用户指南 中的授予最小权限。
- 为敏感操作启用 MFA – 为增强安全性，要求 IAM 用户使用多重身份验证 (MFA) 来访问敏感资源或 API 操作。有关更多信息，请参阅 IAM 用户指南 中的在 AWS 中使用多重身份验证 (MFA)。
- 使用策略条件来增强安全性 – 在切实可行的范围内，定义基于身份的策略在哪些情况下允许访问资源。例如，您可编写条件来指定请求必须来自允许的 IP 地址范围。您也可以编写条件，以便仅允许指定日期或时间范围内的请求，或者要求使用 SSL 或 MFA。有关更多信息，请参阅 IAM 用户指南 中的 IAM JSON 策略元素：Condition。

允许用户创建扩展计划

下面显示允许用户创建扩展计划的权限策略示例。

```
{
    "Version": "2012-10-17",
    "Statement": [
```

```
        {
            "Effect": "Allow",
            "Action": [
              "autoscaling-plans:*",
              "cloudwatch:PutMetricAlarm",
              "cloudwatch:DeleteAlarms",
              "cloudwatch:DescribeAlarms",
              "cloudformation:ListStackResources"
            ],
            "Resource": "*"
        }
    ]
}
```

要使某个用户能够使用扩展计划，该用户必须拥有额外的权限，以允许他们使用其 AWS 账户中的特定资源。这些权限在 其他所需的权限 (p. 23)中列出。

每个控制台用户还需要一些权限，以便他们能够发现其 AWS 账户中的可扩展资源，并从 AWS Auto Scaling 控制台查看 CloudWatch 指标数据的图表。下面列出了使用 AWS Auto Scaling 控制台所需的一组附加的权限：

- `cloudformation:ListStacks`：列出堆栈。
- `tag:GetTagKeys`：查找包含特定标签键的可扩展资源。
- `tag:GetTagValues`：查找包含特定标签值的资源。
- `autoscaling:DescribeTags`：查找包含特定标签的 Auto Scaling 组。
- `cloudwatch:GetMetricData`：查看 AWS 资源的指标图表中的数据。

允许用户启用预测式扩展

下面显示允许用户启用预测式扩展的权限策略示例。这些权限可扩展设置为扩展 Auto Scaling 组的扩展计划的功能。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
              "cloudwatch:GetMetricData",
              "autoscaling:DescribeAutoScalingGroups",
              "autoscaling:DescribeScheduledActions",
              "autoscaling:BatchPutScheduledUpdateGroupAction",
              "autoscaling:BatchDeleteScheduledAction"
            ],
            "Resource": "*"
        }
    ]
}
```

其他所需的权限

授予对 AWS Auto Scaling 的权限时，您必须决定用户要获得针对哪些资源的权限。根据要支持的方案，您可以在 IAM 策略语句的 `Action` 元素中指定以下操作。

Auto Scaling组

要将 Auto Scaling 组添加到扩展计划，用户必须具有来自 Amazon EC2 Auto Scaling 的以下权限：

- `autoscaling:UpdateAutoScalingGroup`

- autoscaling:DescribeAutoScalingGroups
- autoscaling:PutScalingPolicy
- autoscaling:DescribePolicies
- autoscaling:DeletePolicy

ECS 服务

要将 ECS 服务添加到扩展计划，用户必须具有来自 Amazon ECS 和 Application Auto Scaling 的以下权限：

- ecs:DescribeServices
- ecs:UpdateService
- application-autoscaling:RegisterScalableTarget
- application-autoscaling:DescribeScalableTargets
- application-autoscaling:DeregisterScalableTarget
- application-autoscaling:PutScalingPolicy
- application-autoscaling:DescribeScalingPolicies
- application-autoscaling:DeleteScalingPolicy

Spot 队列

要将 Spot 队列添加到扩展计划，用户必须具有来自 Amazon EC2 和 Application Auto Scaling 的以下权限：

- ec2:DescribeSpotFleetRequests
- ec2:ModifySpotFleetRequest
- application-autoscaling:RegisterScalableTarget
- application-autoscaling:DescribeScalableTargets
- application-autoscaling:DeregisterScalableTarget
- application-autoscaling:PutScalingPolicy
- application-autoscaling:DescribeScalingPolicies
- application-autoscaling:DeleteScalingPolicy

DynamoDB 表或全局索引

要将 DynamoDB 表或全局索引添加到扩展计划，用户必须具有来自 DynamoDB 和 Application Auto Scaling 的以下权限：

- dynamodb:DescribeTable
- dynamodb:UpdateTable
- application-autoscaling:RegisterScalableTarget
- application-autoscaling:DescribeScalableTargets
- application-autoscaling:DeregisterScalableTarget
- application-autoscaling:PutScalingPolicy
- application-autoscaling:DescribeScalingPolicies
- application-autoscaling:DeleteScalingPolicy

Aurora 数据库集群

要将 Aurora 数据库集群添加到扩展计划，用户必须具有来自 Amazon Aurora 和 Application Auto Scaling 的以下权限：

- rds:AddTagsToResource
- rds:CreateDBInstance
- rds:DeleteDBInstance
- rds:DescribeDBClusters
- rds:DescribeDBInstances
- application-autoscaling:RegisterScalableTarget
- application-autoscaling:DescribeScalableTargets
- application-autoscaling:DeregisterScalableTarget
- application-autoscaling:PutScalingPolicy
- application-autoscaling:DescribeScalingPolicies
- application-autoscaling:DeleteScalingPolicy

创建服务相关角色所需的权限

当 AWS 账户中的任何用户首次创建启用预测式扩展计划的扩展计划时，AWS Auto Scaling 需要具有创建服务相关角色的权限。如果服务相关角色尚不存在，AWS Auto Scaling 会在您的账户中创建此角色。此服务相关角色向 AWS Auto Scaling 授予权限，以便它能代表您调用其他服务。

为使自动角色创建操作成功，用户必须具有 iam:CreateServiceLinkedRole 操作的权限。

```
"Action": "iam:CreateServiceLinkedRole"
```

下面显示了允许用户创建 AWS Auto Scaling 服务相关角色的权限策略示例。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "arn:aws-cn:iam::*:role/aws-service-role/autoscaling-
plans.amazonaws.com/AWSServiceRoleForAutoScalingPlans_EC2AutoScaling",
            "Condition": {
                "StringLike": {
                    "iam:AWSServiceName":"autoscaling-plans.amazonaws.com"
                }
            }
        }
    ]
}
```

有关更多信息，请参阅AWS Auto Scaling 服务相关角色 (p. 20)。

AWS Auto Scaling 中的日志记录和监控

监控是保持 AWS Auto Scaling 和您的其他 AWS 解决方案的可靠性、可用性和性能的重要环节。您应从 AWS 解决方案的各个部分收集监控数据，以便更轻松地调试出现的多点故障。AWS 提供以下工具来记录和监控活动，并在适当时采取自动措施：

Amazon CloudWatch 警报

CloudWatch 通过自动监控 AWS 资源的特定指标，帮助您检测不正常的应用程序行为。您可以配置 CloudWatch 警报并设置 Amazon SNS 通知，以在指标值不符合预期或检测到特定异常时发送电子

邮件。例如，您可以在网络活动突然高于或低于指标的预期值时收到通知。有关更多信息，请参阅 Amazon CloudWatch 用户指南。

AWS CloudTrail

AWS CloudTrail 捕获由某个 AWS 账户发出或代表该账户发出的 API 调用和相关事件。然后，它将日志文件传送到您指定的 Amazon S3 存储桶。您可以标识哪些用户和账户调用了 AWS、从中发出调用的源 IP 地址以及调用的发生时间。有关更多信息，请参阅 AWS CloudTrail User Guide。有关由 AWS Auto Scaling 记录的 CloudTrail API 调用的信息，请参阅使用 AWS Auto Scaling 记录 CloudTrail API 调用。

Amazon CloudWatch Logs

Amazon CloudWatch Logs 使您能够监控、存储和访问来自 Amazon EC2 实例、CloudTrail 和其他来源的日志文件。CloudWatch Logs 可以监控日志文件中的信息，并在达到特定阈值时通知您。您还可以在高持久性存储中检索您的日志数据。有关更多信息，请参阅 Amazon CloudWatch Logs User Guide。

Amazon CloudWatch Events

CloudWatch Events 提供近乎实时的系统事件流，这些系统事件描述 AWS 资源的变化。AWS Auto Scaling 目标不发出事件。但是，您可以编写规则，用于根据 AWS Auto Scaling 对 API 的调用结果在其他 AWS 服务中触发自动操作。有关更多信息，请参阅 用户指南CloudWatch Events 中的AWS CloudTrail创建使用 对 AWS API 调用触发的 规则Amazon CloudWatch Events。

相关主题

- 中的监控 Auto Scaling 实例和组 Amazon EC2 Auto Scaling 用户指南
- Application Auto Scaling 中的 监控 Application Auto Scaling 用户指南

AWS Auto Scaling 的合规性验证

作为多个 AWS 合规性计划的一部分，第三方审计员将评估 Amazon Web Services (AWS) 服务的安全性和合规性。其中包括 SOC、PCI、FedRAMP、HIPAA 等。

有关特定合规性计划范围内的 AWS 服务列表，请参阅合规性计划范围内的 AWS 服务。有关常规信息，请参阅 AWS 合规性计划。

您可以使用 AWS Artifact 下载第三方审计报告。有关更多信息，请参阅下载 AWS Artifact 中的报告。

您在使用 AWS Auto Scaling 时的合规性责任由您数据的敏感性、贵公司的合规性目标以及适用的法律法规决定。AWS 提供以下资源来帮助满足合规性：

- 安全性与合规性快速入门指南安全性与合规性快速入门指南 – 这些部署指南讨论了架构注意事项，并提供了在 AWS 上部署基于安全性和合规性的基准环境的步骤。
- 《设计符合 HIPAA 安全性和合规性要求的架构》白皮书 – 此白皮书介绍公司如何使用 AWS 创建符合 HIPAA 标准的应用程序。
- AWS 合规性资源 – 此业务手册和指南集合可能适用于您的行业和位置。
- 开发人员指南 中的使用规则评估资源AWS Config 服务评估您的资源配置对内部实践、行业指南和法规的遵循情况。–AWS Config
- AWS Security Hub – 此 AWS 服务提供了 AWS 中安全状态的全面视图，可帮助您检查是否符合安全行业标准和最佳实践。

AWS Auto Scaling 中的弹性

AWS 全球基础设施围绕 AWS 区域和可用区构建。

AWS 区域提供多个在物理上独立且隔离的可用区，这些可用区通过延迟低、吞吐量高且冗余性高的网络连接在一起。

利用可用区，您可以设计和操作在可用区之间无中断地自动实现故障转移的应用程序和数据库。与传统的单个或多个数据中心基础设施相比，可用区具有更高的可用性、容错性和可扩展性。

有关 AWS 区域和可用区的更多信息，请参阅 AWS 全球基础设施。

AWS Auto Scaling 中的基础设施安全性

作为一项托管服务，AWS Auto Scaling 由 AWS 中所述的 Amazon Web Services 全球网络安全程序提供保护： 安全流程概述白皮书。

您可以使用 AWS 发布的 API 调用通过网络访问 AWS Auto Scaling。客户端必须支持传输层安全性 (TLS) 1.0 或更高版本。建议使用 TLS 1.2 或更高版本。客户端还必须支持具有完全向前保密 (PFS) 的密码套件，例如 Ephemeral Diffie-Hellman (DHE) 或 Elliptic Curve Ephemeral Diffie-Hellman (ECDHE)。大多数现代系统（如 Java 7 及更高版本）都支持这些模式。

此外，必须使用访问密钥 ID 和与 IAM 主体关联的秘密访问密钥来对请求进行签名。或者，您可以使用 AWS Security Token Service (AWS STS) 生成临时安全凭证来对请求进行签名。

AWS Auto Scaling 和接口 VPC 终端节点

您可以通过创建接口 VPC 终端节点来在 Virtual Private Cloud (VPC) 与 AWS Auto Scaling API 之间建立专用连接。您可以使用此连接从 VPC 调用 AWS Auto Scaling API，而无需通过 Internet 发送流量。终端节点提供与 AWS Auto Scaling API 之间的可靠、可扩展的连接。它无需 Internet 网关、NAT 实例或 VPN 连接即可完成此操作。

接口 VPC 终端节点由 AWS PrivateLink 提供支持，此功能使用私有 IP 地址在 AWS 服务之间实现私有通信。有关更多信息，请参阅 AWS PrivateLink 。

创建接口 VPC 终端节点

Note

目前AWS Auto Scaling，此功能不适用于 AWS 中国（北京）区域和 AWS 中国（宁夏）区域的 。

您可以使用 AWS Auto Scaling 控制台或 Amazon VPC (AWS Command Line Interface () 为 AWS CLI). 服务创建 VPC 终端节点。使用以下服务名称为 AWS Auto Scaling 创建终端节点：

- com.amazonaws.*region*.autoscaling-plans 为 — API 操作AWS Auto Scaling创建终端节点。

有关更多信息，请参阅 Amazon VPC 用户指南 中的创建接口终端节点.

为终端节点启用私有 DNS，以便使用其默认 DNS 主机名（例如 ）向受支持的服务发出 API 请求。`autoscaling-plans.us-east-1.amazonaws.com`). 为 AWS 服务创建终端节点时，默认情况下会启用此设置。有关更多信息，请参阅 Amazon VPC 用户指南中的通过接口终端节点访问服务.

您不需要更改任何AWS Auto Scaling设置。 使用服务终端节点或私有接口 VPC 终端节点（二者中在使用中的那个AWS Auto Scaling）调用其他 AWS 服务。

创建 VPC 终端节点策略

您可以向 VPC 终端节点附加策略来控制对 AWS Auto Scaling API 的访问。该策略指定：

- 可执行操作的委托人。
- 可执行的操作。
- 可对其执行操作的资源。

以下示例显示了一个 VPC 终端节点策略，该策略拒绝所有人通过终端节点删除扩展计划的权限。示例策略还授予所有人执行所有其他操作的权限。

```
{
   "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "autoscaling-plans:DeleteScalingPlan",
            "Effect": "Deny",
            "Resource": "*",
            "Principal": "*"
        }
    ]
}
```

有关更多信息，请参阅 中的使用 VPC Amazon VPC 用户指南终端节点策略。

终端节点迁移

2019 年 11 月 22 日，AWS Auto Scaling 引入了 `autoscaling-plans.region.amazonaws.com` 作为对 AWS Auto Scaling API 的调用的新默认 DNS 主机名和终端节点。新终端节点与最新版本的 AWS CLI 和 兼容SDKs。 如果您尚未执行此操作，请安装最新的 AWS CLI 和 SDKs 以使用新的终端节点。要更新 AWS CLI，请参阅 中的使用 pip 安装 AWS AWS Command Line Interface 用户指南CLI。有关 AWS 的信息 SDKs，请参阅用于 Amazon Web Services 的工具。

> **Important**
>
> 为了向后兼容，对 `autoscaling.region.amazonaws.com` API 的调用将继续支持现有 AWS Auto Scaling 终端节点。要将`autoscaling.region.amazonaws.com`终端节点设置为私有接口 VPC 终端节点，请参阅 中的 Amazon EC2 Auto Scaling 和接口 VPC Amazon EC2 Auto Scaling 用户指南终端节点。

使用 CLI 或 AWS API 时要调用的终端节点

对于 AWS Auto Scaling 的当前版本，您对 AWS Auto Scaling API 的调用会自动转到 `autoscaling-plans.region.amazonaws.com` 终端节点，而不是 `autoscaling.region.amazonaws.com`.

通过在每个命令中使用以下参数来指定新的终端节点，可以在 CLI 中调用此终端节点：`--endpoint-url https://autoscaling-plans.region.amazonaws.com`.

您还可以在 CLI 中调用旧终端节点，方法是在每个命令中使用 参数来指定该终端节点，但不建议这样做。`--endpoint-url https://autoscaling.region.amazonaws.com`.

有关SDKs用于调用 的各种 APIs，请参阅相关开发工具包的文档，了解如何将请求定向到特定终端节点。有关更多信息，请参阅用于 Amazon Web Services 的工具.

AWS Auto Scaling 服务配额

您的 AWS 账户具有以下 AWS Auto Scaling 默认配额（以前称为限制）。

要请求提高限制，请使用 Auto Scaling 限制表单。确保在增加配额的请求中指定资源的类型，例如 Amazon EC2 Auto Scaling、Amazon ECS 或 DynamoDB。

每账户每区域的默认配额

Item	默认值
每个资源类型的最大可扩展资源数	配额因资源类型而异。 Amazon DynamoDB：3000 Amazon EC2 Auto Scaling 组：200 所有其他资源类型：500
扩展计划的最大数量	100
每个扩展计划的最大扩展指令数	500
每个扩展指令的最大目标跟踪配置数	10

在扩展工作负载时，请牢记服务配额。例如，当您达到某个服务允许的最大容量单位数时，向外扩展操作将会停止。如果需求下降并且当前容量下降，则 AWS Auto Scaling 会再次向外扩展。为避免再次达到此服务配额限制，您可以请求增加配额限制。对于最大资源容量，每个服务都有各自的默认配额。有关其他 AWS 服务的默认配额的信息，请参阅 https://docs.amazonaws.cn/general/latest/gr/aws-service-information.html 中的服务终端节点和配额Amazon Web Services 一般参考。

AWS 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。

- AWS Auto Scaling – 提供 AWS Auto Scaling 相关信息的主要网页。
- AWS Auto Scaling 常见问题 – 客户提出的 AWS Auto Scaling 问题的答案。
- 开发论坛 AWS Auto Scaling 从社区获取帮助。–
- 标记 组和实例Auto Scaling 获取有关标记 组的信息。–Auto Scaling
- 标记用于 DynamoDB – 获取有关标记 Amazon DynamoDB 表或全局二级索引的信息。
- 标记 资源Amazon RDS 获取有关标记 数据库集群的信息。–Aurora
- 使用标签编辑器 – 获取有关使用标签编辑器的信息，包括标签编辑器支持哪些资源。
- Target tracking scaling policies （目标跟踪扩展策略） 适用于 Amazon EC2 Auto Scaling – 获取有关 Amazon EC2 Auto Scaling 组的目标跟踪扩展策略的信息。
- 目标跟踪扩展策略 所有其他资源的 获取有关 – 之外的资源的目标跟踪扩展策略的信息，例如 Amazon EC2 索引和表以及 DynamoDB 服务。Amazon ECS
- API 和 CLI 参考指南 AWS Auto Scaling 可用于创建、修改和删除 计划的 API 调用和 命令的文档。– AWS CLIAuto Scaling
- 使用 记录 API 调用CloudTrail 获取有关监控对您的账户的 API 调用的信息，包括由 –、命令行工具和其他服务进行的调用。AWS 管理控制台

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 AWS Auto Scaling 文档的重要补充部分。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
新增“安全性”章节 (p. 31)	中的新安全性一章可帮助您了解如何在使用 AWS Auto Scaling 用户指南 时应用责任共担模式责任共担模式。AWS Auto Scaling 作为此更新的一部分，已将用户指南的“身份验证和访问控制”一章替换为一个新的、更实用的部分，即适用于 AWS Auto Scaling 的 Identity and Access management。	March 12, 2020
对 Amazon VPC 终端节点的支持 (p. 31)	您现在可以在 VPC 和 AWS Auto Scaling 之间建立私有连接。有关迁移注意事项和说明，请参阅 AWS Auto Scaling 和接口 VPC 终端节点。	November 22, 2019
支持提高最大容量以超出预测容量，以及指南更改 (p. 31)	添加控制台支持，允许扩展计划按指定的缓冲区值增大高于预测容量的最大容量。有关更多信息，请参阅 https://docs.amazonaws.cn/autoscaling/plans/userguide/gs-specify-custom-settings.html#gs-customize-predictive-scaling 中的预测式扩展设置AWS Auto Scaling 用户指南。此版本还包括入门AWS Auto Scaling教程中的几个重写部分。	March 9, 2019
预测式扩展和增强 (p. 31)	现在，您可以使用预测式扩展来主动扩展您的 Amazon EC2 Auto Scaling 组。此版本还增加了以下支持：替换在扩展计划之外创建的扩展策略（例如来自其他控制台的策略）以及控制是否启用您的计划的动态扩展功能。有关更多信息，请参阅 AWS Auto Scaling 入门。	November 20, 2018
对自定义资源设置的支持 (p. 31)	添加了对每个单独资源或多个资源同时自定义各种设置的支持。有关更多信息，请参阅 AWS Auto Scaling 入门。	October 9, 2018
使用标签作为应用程序源 (p. 31)	此版本增加了对指定一组标签作为应用程序源的支持。	April 23, 2018
新增服务 (p. 31)	AWS Auto Scaling 首次发布。	January 16, 2018

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。