

AWS 白皮书

# Amazon EC2 Spot 实例概览



# Amazon EC2 Spot 实例概览: AWS 白皮书

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

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆或者贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

# Table of Contents

摘要和简介 .....	1
摘要 .....	1
简介 .....	1
何时使用 Spot 实例 .....	2
如何启动 Spot 实例 .....	3
Spot 实例运行方式 .....	4
管理 Spot 实例中断 .....	5
Spot 实例限制 .....	6
Spot 实例最佳实践 .....	7
Spot 与其他 AWS 服务集成 .....	8
Amazon EMR 集成 .....	8
EC2 Auto Scaling 集成 .....	8
Amazon EKS 集成 .....	8
Amazon ECS 集成 .....	8
Amazon ECS 与 AWS Fargate Spot 集成 .....	8
Amazon Batch 集成 .....	9
Amazon SageMaker 集成 .....	9
Amazon Gamelift 集成 .....	9
AWS Elastic Beanstalk 集成 .....	9
总结 .....	10
资源 .....	11
文档历史记录和贡献者 .....	12
文档历史记录 .....	12
贡献者 .....	12

# Amazon EC2 Spot 实例概览

发布日期：2021 年 3 月 5 日 ([文档历史记录和贡献者](#))

## 摘要

本白皮书旨在帮助您实现投资价值最大化，提高预测准确性和成本可预测性，营造主人翁精神和成本透明文化，以及持续度量您的优化状态。

本白皮书简要概述了 Amazon EC2 Spot 实例以及如何有效使用它们的最佳实践。

## 简介

除了[按需实例](#)、[预留实例](#)和 [Savings Plans](#) 之外，第四种 [Amazon Elastic Compute Cloud](#) (Amazon EC2) 定价模式是 [Spot 实例](#)。

与按需定价相比，使用 Spot 实例能够以高达 90% 的折扣使用备用 Amazon EC2 计算容量。这意味着，您可以在相同的预算下显著降低运行应用程序的成本或扩大应用程序的计算容量和吞吐量。按需实例和 Spot 实例的唯一区别在于，当 EC2 需要恢复容量时，它会在提前两分钟发出通知的情况下中断 Spot 实例。

与预留实例或 Savings Plans 不同，Spot 实例不需要承诺即可相比按需定价实现成本节省。但是，因为 EC2 可能会在运行 Spot 实例的容量池（实例类型和可用区的组合）中无容量可用时终止这些实例，因此它们最适合灵活工作负载。

## 何时使用 Spot 实例

您可以将 Spot 实例用于各种具有容错能力的灵活应用程序，示例包括无状态 Web 服务器、API 端点、大数据和分析应用程序、容器化工作负载、CI/CD 高性能和高吞吐量计算 (HPC/HTC)、渲染工作负载以及其他灵活的工作负载。

Spot 实例不适用于不灵活、有状态、非容错或在实例节点之间紧密耦合的工作负载。对于不能容忍目标容量偶尔完全不可用的工作负载，也不建议使用 Spot 实例。强烈建议您不要对这些工作负载使用 Spot 实例或尝试故障转移到按需实例来处理中断。

## 如何启动 Spot 实例

最推荐使用 [Amazon EC2 Auto Scaling](#) 服务启动 Spot 实例，因为它让您能够启动和维护所需的容量，并自动请求资源以替换已中断或手动终止的任何资源。在配置 Auto Scaling 组时，您仅需要根据应用程序需求指定实例类型和所需的容量。有关更多信息，请参阅 Amazon EC2 Auto Scaling 用户指南中的 [Auto Scaling 组](#)。

如果您需要更大的灵活性、已经构建自己的实例启动工作流，或者希望控制实例启动或扩展机制的各个方面，建议您评估能否在即时模式下使用 [EC2 机群](#)，将它作为 EC2 Auto Scaling 的替代方案。此同步 API 让您能够指定实例类型和启动要求列表，并提供比 EC2 [RunInstances](#) API 调用更灵活的功能，来用于启动 Spot 实例或按需实例。

当您使用 AWS 服务运行云工作负载时，还可以使用它们启动 Spot 实例。示例包括 [Amazon EMR](#)、[Amazon EKS](#)、[Amazon ECS](#)、[AWS Batch](#) 和 [AWS Elastic Beanstalk](#)。您还可以使用与 AWS 云集成的第三方工具启动 Spot 实例。

您可以使用基础设施即代码工具 ( [AWS CloudFormation](#)、[AWS CDK](#) ) 或 AWS API、CLI 或 SDK 自动启动 Spot 实例。[Spot 蓝图](#) 提供了一个引导式向导，让您能够为符合 Spot 最佳实践的 AWS CloudFormation 和 Hashicorp Terraform 生成基础设施即代码模板。

## Spot 实例运行方式

Spot 实例的运行过程与其他 EC2 实例完全相同。但是，当 Amazon EC2 需要恢复容量时，这些实例可能会被 EC2 中断。

当 EC2 中断您的 Spot 实例时，它将终止、停止该实例或使其休眠，具体取决于您选择的中断行为。

如果 EC2 在运行时间满一小时之前的第一个小时内中断您的 Spot 实例，则您无需为所用时间付费。但是，如果您停止或终止您的 Spot 实例，则需要为不满一小时的时间付费（与按需或预留实例相同）。有关如何为在不同操作系统上运行的中断的 Spot 实例计费的信息，请参阅 EC2 用户指南中的[为中断的 Spot 实例计费](#)。

每个可用区中每种实例类型的 Spot 价格由 EC2 备用容量的长期供求趋势决定。您按照生效的 Spot 价格进行支付，计费到最接近的秒。

您可以选择性指定您的 Spot 实例的最高价。如果未指定最高价，则最高价默认为按需价格。请注意，您支付的价格绝不会超过 Spot 实例运行时生效的 Spot 价格。建议您不要指定最高价，而是让最高价默认采用按需价格。最高价不会增加您启动 Spot 实例的几率，也不会降低您的 Spot 实例被中断的几率（因为 EC2 仍会在需要恢复容量时中断您的 Spot 实例）。

可用区中实例类型的 Spot 价格可能会随时更改，但通常不会频繁更改。AWS 通过 [DescribeSpotPriceHistory](#) API 以及反映来自 API 的数据的 AWS 管理控制台，发布 Spot 实例的当前 Spot 价格和历史价格。这将有助于您评估 Spot 价格在一段时间内的波动水平和时间。

## 管理 Spot 实例中断

正常处理 Spot 实例中断并最大限度减少对性能或可用性影响的最佳方式是将应用程序设计为具有容错能力。为此，您可以利用 EC2 实例再平衡建议和 Spot 实例中断通知。

EC2 实例再平衡建议是一个信号，会在某个 Spot 实例的中断风险升高时通知您。该信号使您有机会在提前两分钟收到 Spot 实例中断通知时主动管理 Spot 实例。您可以决定将工作负载重新平衡到不处于较高中断风险的新的或现有的 Spot 实例。通过在 EC2 Auto Scaling 组中提供容量再平衡功能，您可以轻松地使用该信号。有关更多信息，请参阅 [Amazon EC2 Auto Scaling 容量再平衡](#)。

Spot 实例中断通知是在 Amazon EC2 中断 Spot 实例前两分钟发出的警告。如果您的工作负载具有“时间灵活性”，那么您可以将 Spot 实例配置为在中断时停止或休眠，而不是终止。Amazon EC2 会在中断时自动停止或休眠您的 Spot 实例，并在我们有可用容量时自动恢复这些实例。

您可以使用 EC2 实例再平衡建议和/或 Spot 实例中断通知，考虑将工作负载设计为具有容错能力，这样您就可以捕获通知并将任务的状态保存到存储（例如，Amazon S3、Amazon EFS 或 Amazon FSx）中，保存来自实例的日志文件（或者连续流式传输它们以采取更具容错性的方法），耗尽来自负载均衡器的连接等。

某些 AWS 和第三方服务已经为您处理 Spot 中断，以减少对应用程序的影响。例如，运行 [具有 Spot 实例的托管节点组](#) 的 Amazon EKS 会在发出针对现有节点的再平衡建议或中断通知时自动启动替换 Kubernetes 节点。



## Spot 实例限制

每个区域每个 AWS 账户运行和请求的 Spot 实例数量都有限制。我们根据正在运行的 Spot 实例为完成未处理的 Spot 实例请求而正在使用或将使用的虚拟中央处理器 (vCPU) 的数量，对 Spot 实例限制进行管理。如果您终止 Spot 实例但不取消 Spot 实例请求，则请求数将计入您的 Spot 实例 vCPU 限制，直到 Amazon EC2 检测到 Spot 实例终止并关闭这些请求。

有六种 Spot 实例限制：

- 所有标准 ( A、C、D、H、I、M、R、T、Z ) Spot 实例请求
- 所有 F Spot 实例请求
- 所有 G Spot 实例请求
- 所有 Inf Spot 实例请求
- 所有 P Spot 实例请求
- 所有 X Spot 实例请求

每种限制指定了一个或多个实例系列的 vCPU 限制。有关不同实例系列、实例代和大小的信息，请参阅 [Amazon EC2 实例类型](#)。

对于 vCPU 限制，您可以按照满足您不断变化的应用程序需求的任意实例类型组合启动时所需的 vCPU 数来使用限制。例如，假设您的“所有标准 Spot 实例请求”限制为 256 个 vCPU，则可以请求 32 个 m5.2xlarge Spot 实例 ( 32 x 8 vCPU ) 或 16 个 c5.4xlarge Spot 实例 ( 16 x 16 vCPU )，或者总共 256 个 vCPU 的任何标准 Spot 实例类型和大小的组合。

有关更多信息，请参阅适用于 Linux 实例的 Amazon EC2 用户指南中的 [监控 Spot 实例限制和使用情况](#) 和 [请求提高 Spot 实例限制](#)。

# Spot 实例最佳实践

您的实例类型要求、预算要求和应用程序设计决定了如何针对您的应用程序应用以下最佳实践：

- 灵活使用实例类型。Spot 实例池是一组具有相同实例类型（如 m5.large）和可用区（如 us-east-1a）的未使用 EC2 实例。您应该灵活地选择请求哪些实例类型，以及可以在哪些可用区中部署工作负载。这为 Spot 提供了更好的机会以查找和分配所需数量的计算容量。例如，如果您要使用 c4、m5 和 m4 系列中的大型实例，则不要仅请求 c5.large。
- 使用容量经过优化的分配策略。EC2 Auto Scaling 组中的分配策略可以帮助您预置目标容量，而无需手动查找具有备用容量的 Spot 实例池。我们建议使用容量经过优化的策略，因为该策略会自动从可用性最高的 Spot 实例池中预置实例。由于 Spot 实例容量源自具有最佳容量的池，因此这会降低中断 Spot 实例的可能性。有关分配策略的更多信息，请参阅 Amazon EC2 Auto Scaling 用户指南中的 [Spot 实例](#)。
- 使用主动容量再平衡。容量再平衡功能可在运行中的 Spot 实例提前两分钟收到中断通知时，主动使用新的 Spot 实例扩展您的 Auto Scaling 组，从而帮助您维护工作负载的可用性。启用容量再平衡功能后，Auto Scaling 将尝试主动替换已收到再平衡建议的 Spot 实例，从而使您有机会将工作负载再平衡到不处于较高中断风险的新 Spot 实例。
- 使用集成的 AWS 服务管理 Spot 实例。其他 AWS 服务与 Spot 集成在一起可降低总体计算成本，而无需管理各个实例或机群。我们建议您为适用的工作负载考虑以下解决方案：Amazon EMR、Amazon ECS、AWS Batch、Amazon EKS、SageMaker、AWS Elastic Beanstalk 和 Amazon GameLift。要了解这些服务的 Spot 最佳实践的更多信息，请参阅 [Amazon EC2 Spot 实例研讨会网站](#)。
- 为 Spot 实例选择正确的现代启动工具。如果某个 AWS 集成服务不适合您的工作负载，并且您仍然需要在构建应用程序时控制 Spot 实例的启动，请使用正确的工具。对于大多数工作负载，您应该使用 EC2 Auto Scaling，因为它为广泛的工作负载（如 ELB 支持的应用程序、容器化工作负载和队列处理作业）提供了更全面的功能集。如果您需要对各个请求进行更多控制，并且正在寻找“仅限启动”工具，请使用即时模式下的 EC2 机群直接替代 RunInstances，它具有更广泛的功能，例如实例类型多样化和分配策略。

# Spot 与其他 AWS 服务集成

Amazon EC2 Spot 实例可与多项 AWS 服务集成。

## Amazon EMR 集成

您可以在 Spot 实例上运行 Amazon EMR 集群，显著降低为分析工作负载处理大量数据的成本。您可以使用 [EMR 实例机群](#) 功能轻松地将 Spot 实例与按需实例和预留实例搭配使用，从而能够运行 EMR 集群。您可以使用 [EMR 分配策略](#) 从可用性最高的容量池中启动 Spot 实例。

## EC2 Auto Scaling 集成

您可以使用 [Amazon EC2 Auto Scaling](#) 组启动和管理 Spot 实例，维护应用程序可用性，使实例类型和购买选项（按需/Spot）选择多样化，以及使用动态、计划和预测性扩展策略扩展您的 Amazon EC2 容量。有关更多信息，请参阅 Amazon EC2 Auto Scaling 用户指南中的 [为具有容错能力的灵活应用程序请求 Spot 实例](#)。

## Amazon EKS 集成

您可以使用 Amazon EKS，通过在 EKS 托管节点组中启动 Spot 实例，对基于 Kubernetes 的工作负载进行成本优化。EKS 托管节点组通过用新启动的实例替换即将中断的 Spot 实例来管理整个 Spot 实例生命周期，以减少 Spot 实例中断（当 EC2 需要更多容量时）时对应用程序性能或可用性造成影响的几率。有关更多信息，请参阅 Amazon EKS 用户指南中的 [托管节点组](#)。

## Amazon ECS 集成

您可以在 Spot 实例上运行 Amazon ECS 集群，以降低运行容器化应用程序的运营成本。Amazon ECS 支持自动耗尽即将中断的 Spot 实例。有关更多信息，请参阅 Amazon Elastic Container Service 开发人员指南中的 [使用 Spot 实例](#)。

## Amazon ECS 与 AWS Fargate Spot 集成

如果您的容器化任务可中断且灵活，您可以选择通过 AWS Fargate Spot 容量提供程序运行 ECS 任务，这意味着您的任务将在无服务器容器平台 AWS Fargate 上运行，并且您将受益于 Fargate Spot 带来的成本节省。有关更多信息，请参阅 Amazon Elastic Container Service 开发人员指南中的 [AWS Fargate 容量提供程序](#)。

## Amazon Batch 集成

[AWS Batch](#) 可在 AWS 上规划、安排和执行您的批量计算工作负载。AWS Batch 还可代表您以动态方式请求 Spot 实例，从而降低运行您的批处理作业的成本。

## Amazon SageMaker 集成

Amazon SageMaker 使您可以使用托管的 Spot 实例轻松训练机器学习模型。与按需实例相比，托管的 Spot 训练可将训练模型的成本降低达 90%。SageMaker 代表您来管理 Spot 中断。有关更多信息，请参阅 Amazon SageMaker 开发人员指南中的 [Amazon SageMaker 中的托管 Spot 训练](#)。

## Amazon GameLift 集成

Amazon GameLift 是一个游戏服务器托管解决方案，可以为多人游戏部署、运营和扩展云服务器。Amazon GameLift 中对 Spot 实例的支持让您有机会显著降低托管成本。在创建托管资源机群时，您可以在按需实例或 Spot 实例之间进行选择。虽然 Spot 实例可能会在提前两分钟收到通知时中断，但 Amazon GameLift 的 FleetIQ 可将中断几率降到最低。有关更多信息，请参阅 Amazon GameLift 开发人员指南中的 [结合使用 Spot 实例与 GameLift](#)。

## AWS Elastic Beanstalk 集成

AWS Elastic Beanstalk 是一项易于使用的服务，用于在熟悉的服务器（例如 Apache、Nginx、Passenger 和 IIS）上部署和扩展使用 Java、.NET、PHP、Node.js、Python、Ruby、GO 和 Docker 开发的 Web 应用程序和服务。您只需上传代码，Elastic Beanstalk 便会自动处理从容量预置、负载均衡、自动扩展到应用程序运行状况监控的部署。您可以在 Elastic Beanstalk 环境中使用 Spot 实例来对 Web 应用程序底层基础设施进行成本优化。有关结合使用 Spot 实例与 Elastic Beanstalk 的信息，请参阅 AWS Elastic Beanstalk 开发人员指南中的 [Spot 实例支持](#)。

## 总结

无论您有灵活的计算需求，还是希望在不增加预算的情况下增加容量，Spot 实例可能都是优化 AWS 成本和/或考虑按规模构建的好方法。通过正确构架工作负载，您可以利用 Spot 实例来满足广泛的需求。有关更多信息，请参阅 [Amazon EC2 Spot 实例](#)。

## 资源

- [AWS 架构中心](#)
- [AWS 白皮书](#)
- [AWS 架构月刊](#)
- [AWS 架构博客](#)
- [“这是我的架构”视频](#)
- [AWS 文档](#)

# 文档历史记录和贡献者

## 文档历史记录

要获得有关此白皮书的更新通知，请订阅 RSS 馈送。

更新-历史记录-更改	更新-历史记录-描述	更新-历史记录-日期
<a href="#">次要更新</a>	调整了页面布局。	2021 年 4 月 30 日
<a href="#">次要更新</a>	内容已更新，以反映当前的最佳实践。为了更好地反映内容，已将白皮书的名字从“大规模利用 Amazon EC2 Spot 实例”更改为“Amazon EC2 Spot 实例概览”。	2021 年 3 月 5 日
<a href="#">次要更新</a>	“Spot 实例限制”已更新。	2021 年 2 月 3 日
<a href="#">初次发布</a>	“大规模利用 Amazon EC2 Spot 实例”已发布。	2018 年 3 月 1 日

### Note

要订阅 RSS 更新，您必须为您正在使用的浏览器启用 RSS 插件。

## 贡献者

以下为对此文档有贡献的个人和组织：

- Amilcar Alfaro，AWS 高级产品营销经理
- Erin Carlson，AWS 营销经理
- Keith Jarrett，AWS 业务拓展部门全球成本优化业务拓展主管
- Ran Sheinberg，AWS 首席解决方案架构师