欢迎学习模块 3：添加存储层。

模块概览

章节

1. 最简单的架构
2. 使用 Amazon S3
3. 在 Amazon S3 中存储数据
4. 将数据移入和移出 Amazon S3
5. 为您的架构选择区域

演示

- Amazon S3 版本控制
- Amazon S3 Transfer Acceleration

实验

- 指导实验：托管静态网站
- 挑战实验：为咖啡馆创建静态网站

知识考核

aws

本模块包括以下章节：

1. 最简单的架构
2. 使用 Amazon S3
3. 在 Amazon S3 中存储数据
4. 将数据移入和移出 Amazon S3
5. 为您的架构选择区域

本模块还包括：

- 一个讲师主导的演示，向您展示 Amazon S3 版本控制功能的工作原理。
- 一个讲师主导的演示，向您展示如何配置 Amazon S3 Transfer Acceleration。
- 一个动手指导实验，详细的分步说明介绍了如何创建 Amazon S3 存储桶并将其配置为托管简单网站。
- 一个动手挑战实验，您将部署一个静态网站来支持咖啡馆场景。仅提供有限的指导，因为所涉及的任务与您在本模块中之前完成的指导实验活动非常相似。

最后，您需要完成一个知识考核，以测试您对本模块中涵盖的关键概念的理解程度。

模块目标

学完本模块后，您应该能够：

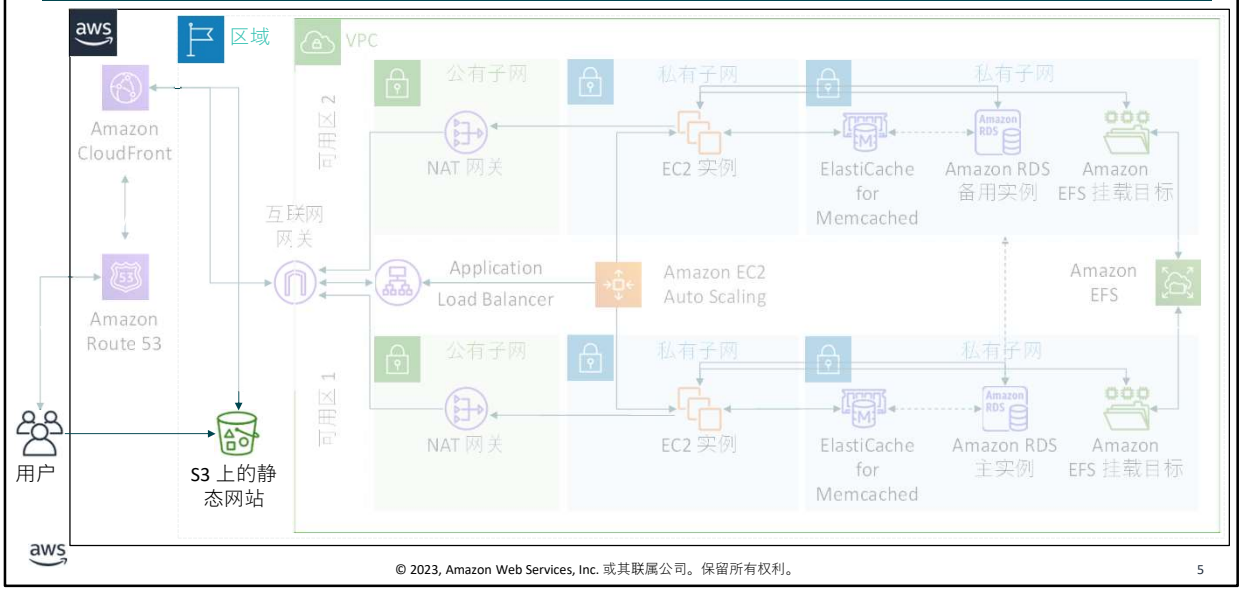
- 识别 Amazon Simple Storage Service (Amazon S3) 可以解决的问题
- 描述如何使用 Amazon S3 高效存储内容
- 了解各种 Amazon S3 存储类和成本注意事项
- 描述如何将数据移入和移出 Amazon S3
- 描述如何选择区域
- 创建静态网站

aws

学完本模块后，您应该能够：
- 识别 Amazon Simple Storage Service (Amazon S3) 可以解决的问题
- 描述如何使用 Amazon S3 高效存储内容
- 了解各种 Amazon S3 存储类和成本注意事项
- 描述如何将数据移入和移出 Amazon S3
- 描述如何选择区域
- 创建静态网站

介绍第 1 节：最简单的架构。

存储是更大架构的一部分



在每个模块介绍新功能时，这张大图中的相应部分便会显示出来。

在本模块中，您首先要学习的是在 AWS 上实施最简单的架构之一，即通过完全在 Amazon S3 上托管该架构来创建静态网站。您还将了解各种 Amazon S3 存储选项以及在 AWS 上选择区域时的一些关键注意事项。

咖啡馆业务需求

咖啡馆刚刚开始运营。他们希望建立一个简单的静态网站，为顾客提供咖啡馆的基本信息（包括菜单、营业时间、位置等）。

这家咖啡馆在这个大城市里只开了一家店，他们在那里出售甜点和咖啡。这家咖啡馆由 Frank 和 Martha 拥有，这是一家夫妻店。他们的女儿 Sofía 和一位名叫 Nikhil 的中学生也在咖啡馆工作。

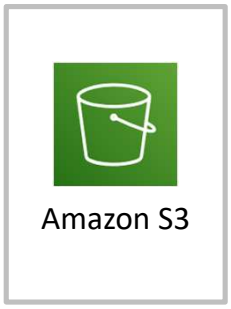
这家咖啡馆目前还没有营销策略。有人路过，注意到咖啡馆，然后决定试一试，这便是他们获得新顾客的唯一方式。咖啡馆供应的优质甜点和咖啡有着不错的口碑，但如果不是熟客或者有人推荐的话，口碑的传播还是比较有限的。

Sofía 建议，他们应该扩大社区对咖啡馆所提供的服务的认识。Frank 和 Martha 同意了。咖啡馆还没有网络服务，他们目前也没有使用任何云计算服务。不过，这一现状即将发生改变。第一个挑战是为咖啡馆创建一个基本的网站。

在本模块中，您将详细了解业务要求以及如何使用 Amazon Web Services 满足这些业务要求。

介绍第 2 节：使用 Amazon S3。

Amazon S3



对象存储服务：

- 存储大量（无限）非结构化数据
- 数据文件作为对象存储在您定义的存储桶中
- 单个对象的最大文件大小为 5 TB
- 所有对象都包含一个 REST 可访问的全局唯一 URL（通用命名空间）
- 所有对象都包含一个键、版本 ID、值、元数据和子资源

Amazon S3 是一项*对象存储服务*。它可让您存储几乎任意数量的数据。数据文件存储为对象。您将对象放置在自己定义的存储桶中。每个存储桶的名称必须在区域之间具有全局唯一性。这意味着存储桶名称在所有 AWS 客户账户中必须是唯一的。

存储的对象的大小可以从 0 字节到 5 TB 不等。虽然单个对象不能大于 5 TB，但是您可以存储的数据总量没有限制。

每个对象都有五个一致的特性。

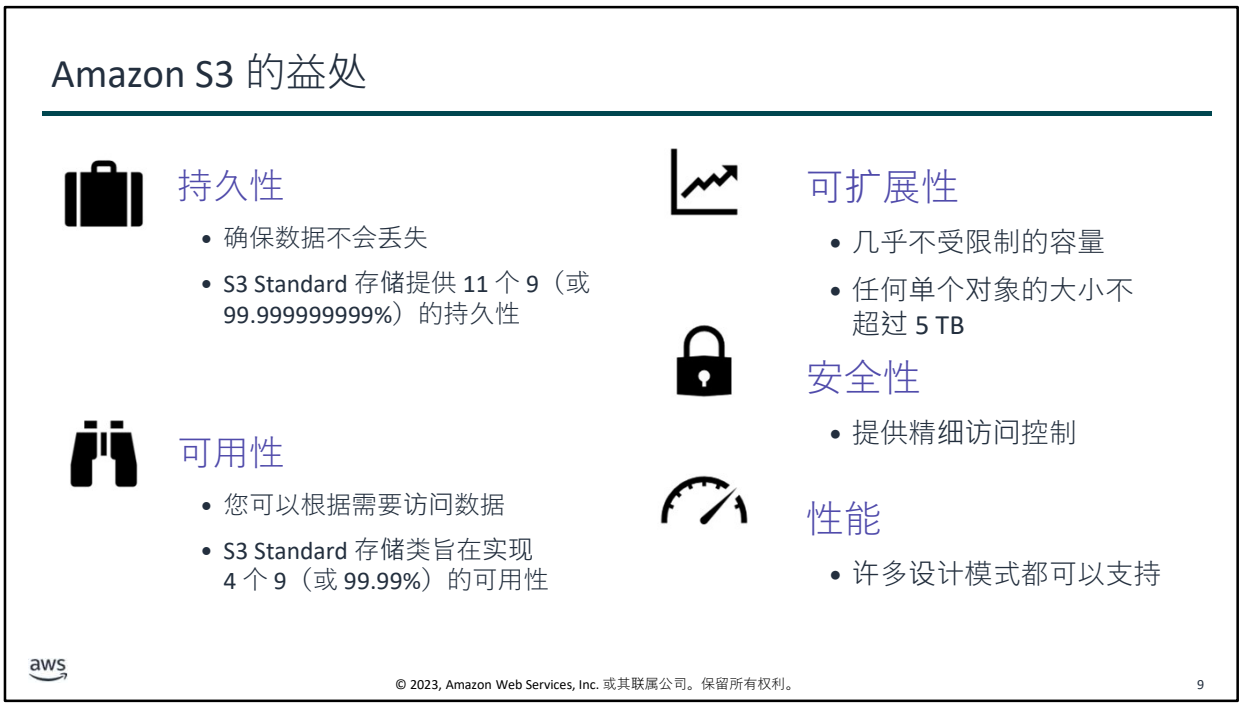
首先，它拥有一个键，这是您分配给对象的名称。您可以使用对象键检索该对象。
在 AWS 管理控制台中，您可以在存储桶内创建目录，然后将对象上传到该目录。但实际上，Amazon S3 并不知道目录，因此键值包括相对于存储桶根目录的完整路径。

对象还包括版本 ID。在存储桶中，键和版本 ID 可以唯一地标识对象。稍后您将在本模块中了解有关版本控制的更多信息。

对象的*值*是您存储的实际内容。它可以是任意序列的字节。对象值是不可变的，这意味着在上传对象之后，您无法修改该值。如果要修改对象，则必须在 Amazon S3 之外进行更改，然后重新上传对象。

对象还包括元数据，它是一组名称值对，可用来存储有关对象的信息。您可以将元数据（称为*用户定义的元数据*）分配给 Amazon S3 中的对象。Amazon S3 也可以将系统元数据分配给这些对象，用于管理对象。

最后，Amazon S3 还使用子资源来存储其他对象特定的信息。



Amazon S3 提供了许多功能，使其成为基于 AWS 构建的许多解决方案的重要组成部分。

首先，它提供*持久性*，以对象平均年度预计损失来衡量。11 个 9 的持久性意味着每年丢失对象的几率为 0.000000001 个百分点。例如，如果您在 Amazon S3 上存储 10,000 个对象，则预期平均每 10,000,000 年发生一次对象丢失。Amazon S3 将您的对象冗余地存储在您所指定的 Amazon S3 区域内多个设施的多台设备上。Amazon S3 可以快速检测和修复任何丢失冗余，从而抵御同时发生的设备故障。Amazon S3 还定期使用校验和来验证您的数据完整性。

Amazon S3 还提供 4 个 9（或 99.99%）的*可用性*。可用性是指您在需要时快速访问数据的能力。它还提供了几乎无限的容量来存储数据，因此它具有*可扩展性*。Amazon S3 具有强大的*安全设置*。它提供了多种方法来控制对所存储数据的访问，还允许您对数据进行加密。

最后，Amazon S3 *具有很高的性能*，对于大多数存储类，第一个字节的延迟以毫秒为单位。有关 S3 性能设计模式的更多信息，请参阅 Amazon S3 文档。常见的方法包括：对频繁访问的内容使用缓存；对在短时间内接收大量请求流量的对象使用可配置的重试和超时逻辑；以及在整个网络中横向扩展和请求并行化以实现高吞吐量。

Amazon S3 常见使用模式

Amazon S3

使用 Amazon S3 可以解决哪些问题？
现在，您将考虑一些使用案例。

既然您已经了解了 Amazon S3 的许多功能，那么，如何使用这些功能来满足您的需求？

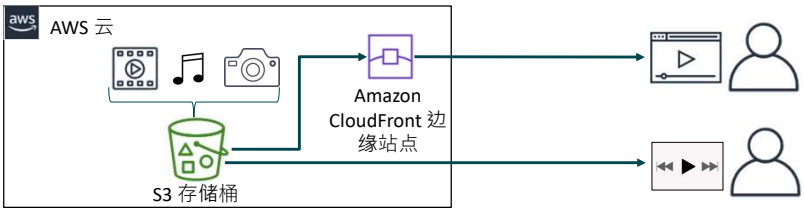
在本模块的这节内容中，您将了解四种使用 Amazon S3 作为强大架构解决方案重要组成部分的常见使用案例。

Amazon S3 使用案例 1：
存储和分发 Web 内容和媒体

构建冗余、可扩展且高度可用的基础设施，以便托管要上传和下载的视频、照片或音乐。

https://<bucket-name>.s3.amazonaws.com

https://<bucket-name>.s3.amazonaws.com/video.mp4



Amazon S3 的常见使用场景是将其用于*媒体托管*。在此使用案例中，Amazon S3 用于存储和分发视频、照片、音乐文件和其他媒体。此内容可以直接从 Amazon S3 交付，因为 Amazon S3 中的每个对象都有唯一的 HTTP URL。

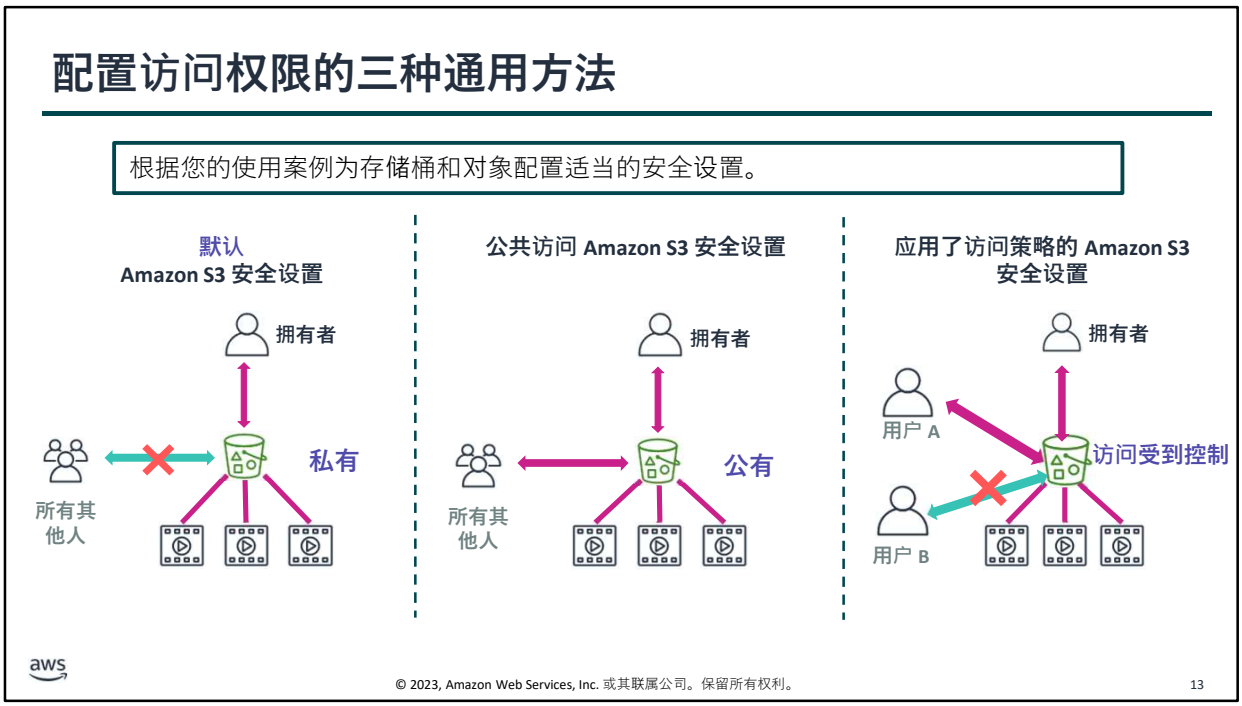
或者，Amazon S3 可以作为内容分发网络 (CDN) 的源存储，例如 *Amazon CloudFront*。Amazon S3 的弹性使其非常适合托管需要带宽来应对极端需求峰值的 Web 内容。此外，由于您不需要为 Amazon S3 预置存储，因此它非常适合托管用户生成的数据密集型内容（如视频和照片共享站点）的快速增长的网站。

保护 Amazon S3 存储桶和对象

- 默认情况下，新创建的 S3 存储桶和对象均为私有并受保护
- 当使用案例必须共享 Amazon S3 数据时–
  - 管理和控制数据访问
  - 遵循最低权限原则
- 用于控制 Amazon S3 数据访问的工具和选项包括–
  - 阻止公共访问功能：在新存储桶上默认启用，易于管理
  - IAM 策略：用户可以使用 IAM 进行身份验证时的理想之选
  - 存储桶策略：您可以定义对特定对象或存储桶的访问
  - 访问控制列表 (ACL)：传统访问控制机制
  - S3 接入点：您可以使用特定于每个应用程序的名称和权限来配置访问
  - 预签名 URL：您可以通过临时 URL 向其他人授予有时限的访问权限
  - AWS Trusted Advisor 存储桶权限检查：免费功能

默认情况下，所有 S3 存储桶都是私有的，只能由获得显式访问授权的用户访问。管理和控制对 Amazon S3 数据的访问至关重要。AWS 提供了许多工具和选项，用于控制对 S3 存储桶或对象的访问，例如：

- 使用 Amazon S3 阻止公共访问。这些设置会覆盖任何其他策略或对象权限。为您不希望提供公开访问权限的所有存储桶启用阻止公共访问。此功能提供了一个直接方法来避免意外泄露 Amazon S3 数据。
- 编写 AWS Identity and Access Management (IAM) 用户策略，以指定可以访问特定存储桶和对象的用户或角色。
- 编写存储桶策略，以确定对特定存储桶或对象的访问权限。此选项通常在用户或系统无法使用 IAM 进行身份验证时使用。存储桶策略可以配置为授予跨 AWS 账户的访问权限，或授予对 Amazon S3 数据的公开或匿名访问权限。如果使用存储桶策略，应对其进行仔细编写并全面测试。您可以在存储桶策略中指定一个拒绝语句来限制访问。即使用户具有附加到用户、基于身份的策略中授予的权限，访问也将受到限制。
- 创建 S3 接入点。接入点是唯一的主机名，对通过它发出的请求强制执行不同的权限和网络控制。拥有共享数据集的客户可以通过创建个性化的接入点，并为每个应用程序定制名称和权限，从而扩展对许多应用程序的访问。
- 针对存储桶和对象设置访问控制列表 (ACL)。ACL 并不常用（ACL 早于 IAM）。如果您使用 ACL，请勿将访问权限设置得太开放或太宽松。
- AWS Trusted Advisor 提供存储桶权限检查功能。这是一款非常实用的工具，可用于发现账户中是否有任何存储桶具有授予全局访问的权限。



以下是配置 S3 存储桶中对象访问权限的三种不同的常见方法。

*左侧的场景*显示了 Amazon S3 的默认安全设置。默认情况下，所有 Amazon S3 存储桶及其中存储的对象都是私有（*受保护*）的。只有账户管理员和 AWS 根用户才有权访问新创建且未经修改的存储桶。 资源拥有者可以向其他人授予特定的访问权限，但未被授予这些权限的任何人都将无法访问。

*中间的场景*显示了 S3 安全设置已被禁用，任何人都可以公开访问存储桶中存储的对象的情况。

注意！使用 Amazon S3 存储桶托管静态网站是快速设置 AWS 架构的一个示例。但是，对于大多数 Amazon S3 使用案例，您不想授予对 Amazon S3 的公共访问权限。 大多数使用案例不需要公共访问权限。 更常见的情况是，您在 Amazon S3 之外运行应用程序，而使用 Amazon S3 来存储该应用程序所使用的数据，或备份敏感数据。对于这些常见的使用案例，永远不应该授予对存储数据的存储桶的公共访问权限。

*右侧的场景*显示了 Amazon S3 配置为提供*受控访问*的情况。用户 A 被授予了对存储桶中对象的访问权限，但用户 B 被拒绝访问。受控访问场景很常见。存储桶拥有者可以通过本模块前面讨论过的一个或多个工具或选项来配置这些场景，控制对 Amazon S3 数据的访问权限。

考虑在 Amazon S3 中加密对象

- 加密将使用密钥对数据进行编码，使其不可读
  - 只有具有密钥的用户才可对数据解码
  - 或者，使用 AWS Key Management Service (AWS KMS) 来管理密钥

- *服务器端*加密
  - 在存储桶上，通过选择 Default encryption（默认加密）选项启用此功能
  - Amazon S3 会在将对象保存到磁盘之前加密对象，并在您下载对象时解密对象

- *客户端*加密
  - 在客户端加密数据并将加密的数据上传到 Amazon S3
  - 在这种情况下，您负责管理加密过程

当您的目标是保护数字数据时，数据加密是必不可少的工具。数据加密是对可读数据进行编码。如果用户没有权限访问用于解码加密数据的密钥，则无法访问相关数据。因此，即使攻击者获得了对数据的访问权限，他们也无法获取有价值的信息。

对于存储在 Amazon S3 中的数据，您有两种主要的加密选项。

当您在存储桶上设置默认加密选项时，它会启用服务器端加密。使用此功能，Amazon S3 会在将对象保存到磁盘之前加密该对象。然后，Amazon S3 将在您下载对象时对其进行解密。

另一个选项是客户端加密。使用此方法时，在您将数据上传到 Amazon S3 之前，您需要先在客户端对数据进行加密。在这种情况下，您需要管理加密过程、加密密钥和相关工具。如服务器端加密一样，客户端加密可以帮助减少面临的风险：通过使用存储在一个不同机制（而不是存储数据本身的机制）中的密钥来加密数据。

Amazon S3 使用案例 2：托管静态网站

第二个 Amazon S3 使用案例是使用该服务托管静态网站。在静态网站上，各个网页都包含静态内容，还可能包含客户端脚本。

相比之下，*动态*网站依赖服务器端处理，这可能涉及为响应服务器端脚本（例如 PHP, JSP 或 ASP.NET）而运行的数据库查询。Amazon S3 不支持服务器端脚本编写。但是，AWS 还提供其他服务，使您能够托管动态网站。

要托管*静态*网站，请为网站托管配置 S3 存储桶。然后，将您的网站内容上传到存储桶。

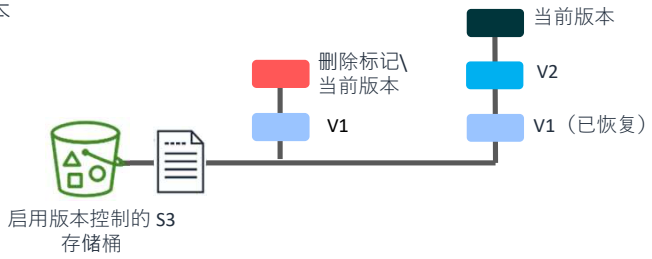
该示例显示静态站点可能包含 HTML 文件、图像、视频和 JavaScript 等格式的客户端脚本。

使用这种方法，您无需运行托管 Web 服务器的虚拟机。事实上，您不需要运行服务器。但是，您仍然可以托管网站。Amazon S3 推出了低成本的 Web 托管解决方案，其中包括高性能、可扩展性和可用性。

Amazon S3 最佳实践：版本控制

- 防止意外覆盖和删除，不会造成任何性能损失
- 每次上传生成一个新版本
- 允许轻松检索已删除的对象或回滚到早期版本
- S3 存储桶的三种可能状态–

1. *默认：未启用版本控制*
2. 启用版本控制
3. 暂停版本控制



Amazon S3 为客户提供具有高安全性和持久性的存储基础设施。版本控制进一步提高了保护等级。它提供了一种在应用程序发生故障或客户意外覆盖或删除对象时恢复数据的方法。

*版本控制*是在相同的存储桶中保留对象的多个变体的方法。对于 S3 存储桶中存储的每个对象，您可以使用版本控制功能来保留、检索和还原它们的各个版本。

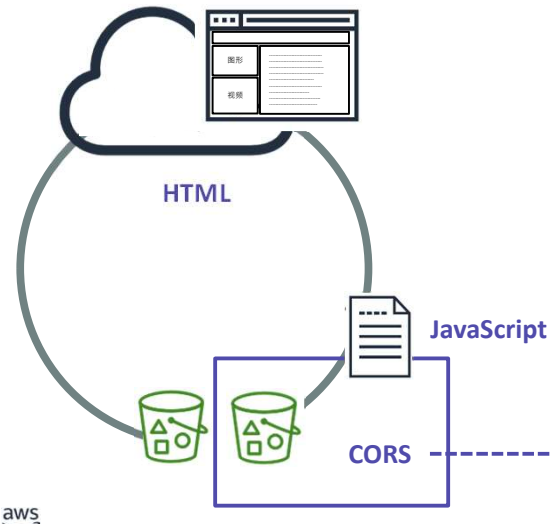
- 如果删除（而不是永久移除）对象，Amazon S3 会插入一个删除标记，该标记将成为当前对象的版本。您始终可以还原以前的版本。
- 覆盖对象会导致在存储桶中产生一个新的对象版本。您始终可以还原以前的版本。

存储桶可处于以下三种状态之一：未启用版本控制（默认）、启用版本控制或暂停版本控制。为存储桶启用版本控制后，将无法将其更改为未启用版本控制状态。但是，您可以暂停该存储桶的版本控制。

演示：Amazon S3 版本控制

现在，讲师可能会选择使用 AWS 管理控制台演示 Amazon S3 版本控制。

支持跨源资源共享 (CORS)



示例 CORS 配置

```
<CORSConfiguration>
 <CORSRule>

<AllowedOrigin>http://www.example.com</AllowedOrigin>
 <AllowedMethod>PUT</AllowedMethod>
 <AllowedMethod>POST</AllowedMethod>
 <AllowedMethod>DELETE</AllowedMethod>
 <AllowedHeader>*</AllowedHeader>
 <MaxAgeSeconds>3000</MaxAgeSeconds>
 <ExposeHeader>x-amz-server-side-
encryption</ExposeHeader>
 <ExposeHeader>x-amz-request-id</ExposeHeader>
 <ExposeHeader>x-amz-id-2</ExposeHeader>
 </CORSRule>
</CORSConfiguration>
```

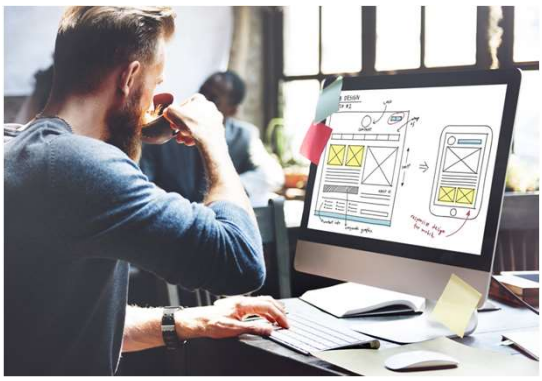

跨源资源共享 (CORS) 为在一个域中加载的客户端 Web 应用程序定义了一种与另一个域中的资源进行交互的方法。借助 CORS 支持，您可以使用 Amazon S3 构建丰富的客户端 Web 应用程序，并有选择地允许跨源访问 Amazon S3 资源。

要将您的存储桶配置为允许跨源请求，您可以创建 CORS 配置。CORS 配置是一个 XML 文档，其中包含识别以下内容的规则：

- 允许访问您的存储桶的源。
- 每个源支持的操作（HTTP 方法）。在本例中，PUT、POST 和 DELETE 请求来自 http://www.example.com，可以使用 Amazon Route 53 将其配置到另一个 S3 存储桶。
- 其他特定于操作的信息。

有关 CORS 的更多信息，请参阅跨源资源共享 (CORS) AWS 文档。

模块 3 – 指导实验：托管静态网站



您现在将完成模块 3 – 指导实验：托管静态网站。

指导实验：任务

1. 在 Amazon S3 中创建存储桶
2. 将内容上传到存储桶
3. 允许访问对象
4. 更新网站

在本指导实验中，您将完成以下任务：

1. 在 Amazon S3 中创建存储桶
2. 将内容上传到存储桶
3. 允许访问对象
4. 更新网站

大约 20 分钟

现在可以开始指导实验了。

指导实验总结：要点

完成这个指导实验之后，您的讲师可能会带您讨论此指导实验的要点。

Amazon S3 使用案例 3：
用于计算和分析的数据存储

您还可以将 Amazon S3 用作数据存储进行计算或大规模分析，如金融交易分析、点击流分析和媒体转码。Amazon S3 可以支持这些工作负载，因为它可以横向扩展，从而实现多个并发事务。

在此示例中，当 Spot 实例的出价较低时，或当 Amazon EMR 集群启动时，Amazon Elastic Compute Cloud (Amazon EC2) Spot 实例集就会启动。无论如何，在计算容量可用之后，将从 Amazon S3 以及从另一个数据源中提取*未经处理的原始*数据。数据通过集成和转换数据的计算算法运行。生成的*处理*数据将加载到其他 Amazon S3 存储桶中。现在数据已处理完毕，计算容量将终止以节省成本。最后，可能会使用 Amazon QuickSight 等分析工具从处理的数据中获取有意义的见解。这只是 Amazon S3 如何在大规模分析解决方案架构中为数据存储发挥重要作用的示例场景之一。

Amazon S3 使用案例 4：备份和归档关键数据

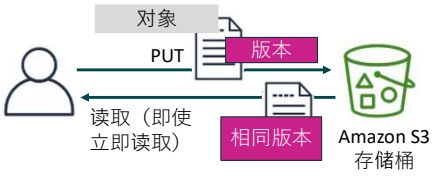
在本模块讨论的第四个也是最后一个使用案例中，Amazon S3 被用作数据备份解决方案。由于具备高持久性和可扩展性，Amazon S3 适合作为数据备份和归档工具。

在这种情况下，数据是从本地部署企业数据中心以及大量 Amazon EC2 服务器备份的。这些服务器运行生成数据的应用程序。

此外，您可以将长期数据从 Amazon S3 Standard 存储移动到 Amazon Simple Storage Service Glacier。本模块稍后将进一步详细讨论此过程。您可以在存储桶上配置另一个 Amazon S3 选项以实现更高的持久性，这就是*跨区域复制*。在跨区域复制中，上传到一个区域中的存储桶的对象将自动复制到其他区域中的其他 S3 存储桶。

Amazon S3 数据一致性模型

- Amazon S3 对所有区域中的所有新的和现有的*对象*都具有**强一致性**
  - 为 S3 存储桶中对象的所有 GET、LIST 和 PUT 操作提供先写后读一致性
  - 一致性模型为大数据工作负载提供了优势
- 存储桶配置具有最终一致的模型



许多客户开发的大数据分析应用程序都使用 Amazon S3 进行对象存储。这些应用程序通常需要在写操作之后立即访问对象。在 2020 年 12 月之前，Amazon S3 在所有区域为覆盖 PUTS 和 DELETES 提供最终一致性。不过，Amazon S3 现在对所有 AWS 区域的所有新的和现有的 S3 对象都具有很强的一致性。

Amazon S3 通过在 AWS 数据中心的多台服务器之间复制数据来实现高可用性。如果 PUT 请求成功，数据将被安全存储。在 PUT 响应成功后启动的任何读取（GET 或 LIST）都将返回 PUT 写入的数据。这种先写后读强一致性自动存在于所有应用程序中，不会影响性能或可用性。

借助强一致性，无需为支持应用程序而进行更改，从而简化了本地部署分析工作负载的迁移。也无需额外的基础设施（如 S3Guard）来提供强一致性。

虽然对象具有强一致性，但 Amazon S3 存储桶配置具有最终一致性模型。例如，如果删除一个存储桶并立即列出所有存储桶，则已删除的存储桶可能仍会出现在列表中。不过，在短时间内，如果再次运行 list bucket 命令，已删除的存储桶将不再出现在 list buckets 结果中。

有关详细信息，请阅读 Amazon S3 强一致性文档。

第 2 节要点

- 存储桶必须具有全局唯一名称，并在区域级别定义
- 存储桶默认是私有的，处于受保护状态
- Amazon S3 安全性可以通过 IAM 策略、存储桶策略、访问控制列表、S3 接入点和预签名 URL 进行配置
- Amazon S3 对所有区域中的所有新对象和现有对象都具有很强的一致性
- 单个对象的最大大小为 5 TB
- Amazon S3 通常用作计算和分析的数据存储，以及关键数据的备份和归档服务

本模块中这节内容的要点包括：

- 存储桶必须具有全局唯一名称，并在区域级别定义
- 存储桶默认是私有的，处于受保护状态
- Amazon S3 安全性可以通过 IAM 策略、存储桶策略、访问控制列表、S3 接入点和预签名 URL 进行配置
- Amazon S3 对所有区域中的所有新对象和现有对象都具有很强的一致性
- 5 TB 是单个对象的最大大小，但是您可以存储几乎无限的对象
- Amazon S3 通常用作计算和分析的数据存储，以及关键数据的备份和归档服务

介绍第 3 节：在 Amazon S3 中存储数据。

现在您已经使用 Amazon S3 构建了网站，下面是不同 Amazon S3 存储类及其特性的比较。

S3 Standard 为频繁访问的数据提供高持久性、可用性和高性能的对象存储。S3 Standard 提供较低的延迟和较高的吞吐量，因此非常适合各种使用案例，包括云应用程序、动态网站、内容分发、移动和游戏应用程序以及大数据分析。它提供了至少三个可用区的持久性。

S3 标准 - 不频繁访问存储 (S3 Standard-IA) 具有 Amazon S3 Standard 的所有优势，但它在不同的成本模型上运行以存储不经常访问的数据，例如较旧的数字图像或较旧的日志文件。对于放置在其中的任何数据，都需要收取 30 天的最低存储费，而且从 S3 Standard-IA 检索数据的成本也高于从 S3 Standard 存储中检索数据的成本。

S3 One Zone-IA 将数据存储在单个可用区内。它非常适合想要使用较低费用选项并且不需要 S3 Standard 或 S3 Standard-IA 的可用性和弹性的客户。如果用于存储本地部署数据的辅助备份副本或可轻松重新创建的数据，它是一个很好的选择。您还可以将其用作从另一个 AWS 区域复制的数据的经济高效存储。

S3 Intelligent-Tiering 旨在通过自动将数据转移到最经济高效的访问层来优化成本，而不会影响性能，也不会产生运营开销。只需针对每个对象收取小额月度监控和自动化费用，Amazon S3 会监控 S3 Intelligent-Tiering 中对象的访问模式。它将连续 30 天未访问的对象移动到不频繁访问层。如果不频繁访问层中的某个对象被访问，即会自动将该对象移回频繁访问层。在使用 S3 Intelligent-Tiering 时不收取检索费用，并且在各层之间移动对象时不收取额外的分层费用。

Amazon S3 Glacier 是一种安全、持久且成本低的存储类，可用于数据归档。您可以放心存储任意数量的数据，成本与本地部署解决方案相当，甚至更低。为了保持低成本，但适合不同的需求，您有三种检索数据的选择，访问时间和成本各不相同：

- 加速检索通常在 1 到 5 分钟内完成

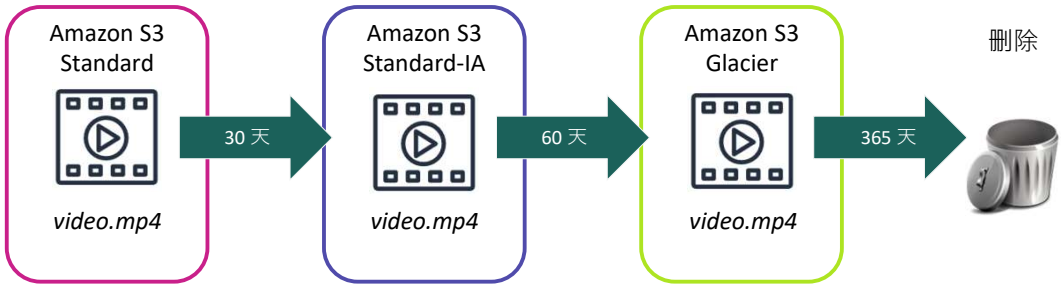
- 标准检索通常在 3 到 5 小时内完成
- 批量检索通常在 5 到 12 小时内完成

Amazon S3 Glacier Deep Archive 是 Amazon S3 最低成本的存储类。它支持长期保留和数字化保存一年中可能被访问一两次的数据。数据至少存储在三个地理位置分散的可用区中，受到 11 个 9（99.999999999%）的持久性保护，可在 12 小时内恢复。

有关 Amazon S3 存储类的更多详细信息，请参阅 AWS 文档。

Amazon S3 **生命周期策略**

配置 **Amazon S3 生命周期策略**可根据时间来删除或移动对象。



您可以配置对象的生命周期，以管理对象在整个生命周期中的存储方式。**生命周期配置**是一组规则，用于定义 Amazon S3 对一组对象应用的操作。

设置 S3 生命周期策略之后，无需更改您的应用程序，*您的数据将自动传输到不同存储类*。

利用生命周期策略，您可以让数据在不同的 Amazon S3 存储类型之间定期循环。这种循环可降低总体成本，因为随着时间的推移，数据的重要性会降低，因此支付的数据费用也会减少。除了可以针对对象设置生命周期规则外，您还可以针对存储桶设置生命周期规则。

有关对象生命周期管理的更多详细信息，请参阅对象生命周期管理 AWS 文档。

Amazon S3 成本

仅按实际用量付费，包括：
存储的对象的 GB（每月）。每个*区域*和每个*存储类*的定价不同。
传出到其他区域或互联网。
PUT、COPY、POST、LIST、GET、SELECT、生命周期转换、数据检索请求。

以下服务免费：
数据从互联网传入到 Amazon S3。
在同一 AWS 区域内的 S3 存储桶之间或从 Amazon S3 传输到同一 AWS 区域内的任何服务。
传出到 Amazon CloudFront。
DELETE 和 CANCEL 请求。

使用 Amazon S3 时，您只需为实际使用量付费。没有最低收费。选择最适合您数据资料的 Amazon S3 存储类时，您需要考虑四项：存储定价、请求和数据检索定价、数据传输和传输加速定价以及数据管理功能定价。

有关 Amazon S3 定价的详细信息，可以在 Amazon S3 定价中找到。

第 3 节要点

- Amazon S3 存储类包括–
  - S3 Standard
  - S3 Standard-IA
  - S3 One Zone-IA
  - S3 Intelligent-Tiering
  - S3 Glacier
  - S3 Glacier Deep Archive
- Amazon S3 生命周期策略可以根据时间删除对象或将对象移动到较便宜的存储类
- 数据传输从互联网传入 Amazon S3 是免费的，但是传输到其他区域或互联网需要付费

本模块中这节内容的要点包括：

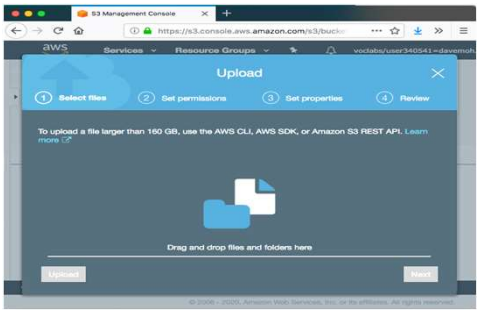
- Amazon S3 存储类包括 – S3 Standard、S3 标准 - 不频繁访问存储、S3 One Zone-不频繁访问存储、S3 Intelligent-Tiering、S3 Glacier 和 S3 Glacier Deep Archive
- Amazon S3 生命周期策略可以根据时间删除对象或将对象移动到较便宜的存储类
- 数据传输从互联网传入 Amazon S3 是免费的，但是传输到其他区域或互联网需要付费

介绍第 4 节：将数据移入和移出 Amazon S3。

将对象移动到 Amazon S3

AWS 管理控制台

使用浏览器上传或下载。



AWS 命令行界面

从终端命令提示符或脚本的调用中上传或下载。

- 上传命令示例：

```
$aws s3 cp test.txt \
s3://AWSDOC-EXAMPLE-BUCKET/test.txt
```

AWS 工具和 SDK

使用 AWS 工具或 SDK 以编程方式移动对象。

在本模块前面的指导实验中，您使用 AWS 管理控制台提供的 Web 浏览器界面将文件上传到 Amazon S3。这是将数据移入或移出 Amazon S3 的最简单方法。它提供了一种基于向导的方法，其中包括将要复制的文件拖放到存储桶中的选项。

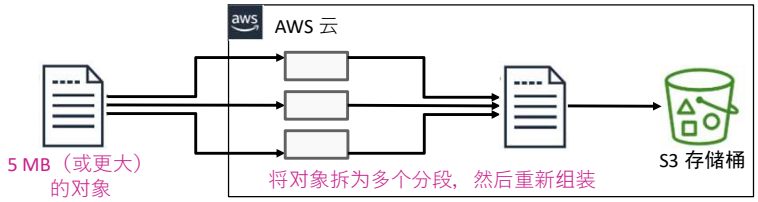
在模块的这节内容中，您将了解一些其他选项，这些选项可用于将数据移入和移出 Amazon S3。

其中两个选项包括使用 AWS 命令行界面 (AWS CLI) 或 AWS SDK。

下面的示例显示了 AWS CLI 上传命令。在命令中，您指定 *aws* 调用 AWS CLI，然后指定服务，即 *S3*。接下来，您发出一个 *cp*（或 copy）子命令，后跟 *test.txt*，这是要复制的本地文件（存在于您的电脑上）。最后，*s3://AWSDOC-EXAMPLE-BUCKET/test.txt* 参数指示应上传文件的存储桶，以及应存储对象值（内容）的键 (*AWSDOC-EXAMPLE-BUCKET/test.txt*)。

S3 AWS CLI 命令参考提供了更多详细信息。

分段上传



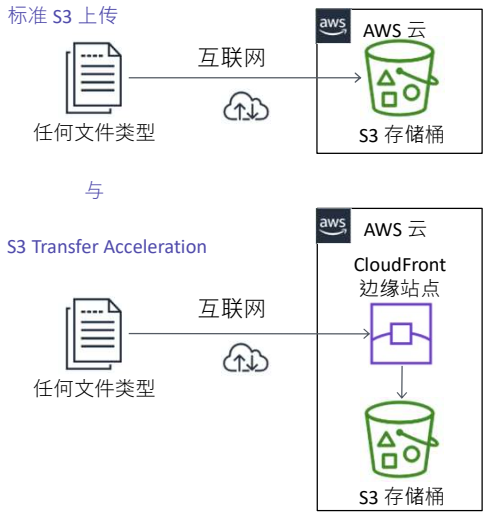
- 文件可以使用分段上传 API 进行上传
  - 您可以将单个对象分段进行上传
  - 每个分段都是对象数据的连续部分
  - 上传完对象的所有分段后，Amazon S3 将汇集这些分段并创建对象
- 通常只用于大于 100 MB 的文件
- 优点–
  - 快速从网络问题中恢复：如果任何分段传输失败，只需重新传输该分段
  - 能够暂停和恢复对象上传
  - 提高吞吐量：并行上传分段以提高吞吐量

分段上传 API 可让您通过可控的分段来持续上传大型对象。优势包括：

- 提高吞吐量 – 您可以并行上传各个分段以提高吞吐量。
- 从任何网络问题快速恢复 – 较小的各个分段可以最大限度地减少因网络错误而重新启动失败上传的影响。
- 暂停和恢复对象上传 – 您可以用一段时间来上传对象分段。启动分段上传后，就不会过期。您必须明确完成或停止分段上传。
- 在您知道对象的最终大小前开始上传 – 您可以在创建对象时就将其上传。
- 上传大型对象 – 您可以使用分段上传 API 来上传高达 5 TB 的大型对象。

请注意，文件的大小必须至少为 5 MB 才能使用分段上传功能。

Amazon S3 Transfer Acceleration



- 加快 Amazon S3 数据传输
- 使用优化的网络协议和 AWS 边缘基础设施
- 典型的速度改善：
  - 跨国传输较大对象的速度可提高 50-500%
  - 在某些情况下可以更高
- Amazon S3 Transfer Acceleration 速度比较工具*
  - 显示获得的速度优势（按区域）

* 有关更多信息，请参阅 http://s3-accelerate-speedtest.s3-accelerate.amazonaws.com/en/accelerate-speed-comparsion.html。

Amazon S3 Transfer Acceleration 利用分布在全球的 Amazon CloudFront 和 AWS 边缘站点，快速、轻松地将数据传输到 S3 存储桶。随后数据将通过经过优化的网络路径路由至 Amazon S3。

适合使用 Transfer Acceleration 的场景：

- 您位于全球各地的客户需要上传到一个集中的存储桶
- 您定期跨大洲传输 GB 或 TB 级的数据
- 通过互联网将文件上传到 Amazon S3 时对可用带宽的利用率不足

演示：
S3 Transfer Acceleration

现在，讲师可能会选择演示 S3 Transfer Acceleration 工具。

将大量数据移动到 Amazon S3：AWS Snowball

AWS Snowball
PB 级数据传输

AWS Snowball

- 可以将数 TB 的数据传入或传出 Amazon S3
  - 可以使用多个设备来传输 PB 级数据
- 解决了对大型数据传输的担忧（网络成本、传输时间、安全性）
  - *示例*：要以 10 Gbps 的上传速度在互联网上传输 10 PB（1000 万 GB）数据，将需要 100 天以上的时间
- 使用方法–
  - 在 AWS 管理控制台中创建一项任务，然后等待 Snowball 设备送达。
  - 连接到本地网络，然后下载并运行 Snowball 客户端
  - 选择要传输（加密）到设备的文件目录
  - 运回设备并跟踪状态

AWS Snowball 是一种 PB 级数据传输方案，您无需编写任何代码或购买任何硬件即可传输数据。您只需在 AWS 管理控制台中创建一个任务，我们就会将 Snowball 设备运送给您。只需将设备接入您的本地网络，然后直接将文件传输到设备上。然后，将设备运回并跟踪货物状态。数据到达安全的 Amazon 设施后，数据将被传输到您的 AWS 账户中。

将大量数据移动到 Amazon S3：AWS Snowmobile

AWS Snowmobile
EB 级数据传输

AWS Snowmobile

- 一个 45 英尺长（13.7 米）的运输集装箱，由半挂卡车牵引
- 每个 Snowmobile 最多可以传输 100 PB
- 提供多层安全保护–
  - 专门的安保人员
  - GPS 跟踪、警报监控、全天候视频监控
  - 运输过程中可选配护送安保车
  - 数据采用 256 位加密密钥进行加密

AWS Snowmobile 是更大的数据传输方案，可传输 EB 级的数据。1 EB 相当于 100 万 TB 或 10 亿 GB。它只用于将极大量的数据迁移到 AWS。Snowmobile 是一种坚固的运输容器，长达 45 英尺（13.7 米），由半挂式卡车牵引。每个 Snowmobile 可以传输 100 PB 数据。

如果尝试在互联网上传输 100 PB 的数据，上传速度为 10 Gbps（假设 TCP/IP 开销为 10%），则需要大约 1018 天（近三年）才能完成数据上传。这不切实际。在这种情况下，使用 AWS Snowmobile 传输数据将是更好的选择。

Snowmobile 采用多层安全防护措施，力保您的数据安全，这些措施包括：专门的安保人员、GPS 跟踪、警报监控、全天候视频监控，并且在运输过程中可选配护送安保车。所有数据都使用 256 位加密密钥进行加密，密钥可通过 AWS Key Management Service (AWS KMS) 管理，这样可确保您的数据安全并形成完整的监管链。

第 4 节要点

- 对于大于 100 MB 的文件以及网络连接可能不稳定的情况，S3 分段上传选项是一个不错的选择
- Amazon S3 Transfer Acceleration 使用边缘站点，可以显著提高上传速度
- AWS Snowball 提供了一种传输 *PB 级*数据的方法，AWS Snowmobile 提供了一种将 *EB 级*数据传输到 AWS 的方法

本模块中这节内容的要点包括：

- 对于大于 100 MB 的文件以及网络连接可能不稳定的情况，S3 分段上传选项是一个不错的选择
- Amazon S3 Transfer Acceleration 使用边缘站点，可以显著提高上传速度
- AWS Snowball 提供了一种传输 PB 级数据的方法，而 AWS Snowmobile 提供了一种将 EB 级数据传输到 AWS 的方法

介绍第 5 节：为您的架构选择区域。

选择区域：合规性和延迟注意事项

- 数据属地和监管合规
  - 是否有相关的区域数据隐私法？
  - 客户数据可以存储在该国家/地区之外吗？
  - 您能否满足监管要求？

- 用户与数据之间的距离
  - 延迟方面的细微差异可能会影响客户体验
  - 选择离用户最近的区域

在决定将数据托管到哪个区域时，需要考虑很多因素。

首先，您应该考虑*数据隐私法*和监管合规要求。您在 AWS 上存储的数据需要遵守数据存储地的国家/地区法律和地方性法规。此外，一些法律规定，如果您在其司法管辖区内经营业务，则不得将相关数据存储到其他地方。与此类似的是，*合规性标准*（例如《美国健康保险流通与责任法案》(HIPAA)）也对数据的存储方式和位置提出了严格要求。

其次，*距离*是在选择区域时的一个重要因素，尤其是当*延迟*对您来说非常重要时，更是如此。大多数情况下，选择最近的区域和选择最远的区域，这两者之间的延迟差异是相对较小的，但即使是细微的延迟差异，也会影响客户体验。客户需要响应迅捷的环境，而且随着时间推移，技术变得越来越强大，客户的这种期望也越来越强烈。

选择区域：服务可用性和成本注意事项

- 服务和功能可用性
  - 并非所有的 AWS 服务都能在任何区域中提供
    - 请参阅 AWS 区域表了解详细信息
    - 服务会定期扩展到新的区域
  - 可以跨区域使用某些服务，但会增加延迟

- 成本效益
  - 成本因区域而异
  - 某些服务（如 Amazon S3）针对传出数据收费
  - 考虑在其他区域内复制整个环境的成本效益

选择区域时，第三个重要的考虑因素是 AWS 服务和功能的可用性。虽然 AWS 一直努力让其服务和功能在所有区域可用，但由于业务遍布全球而导致的复杂性使得实现这个目标非常困难。我们的服务在准备就绪时会先在部分区域发布，然后尽快推向其他区域，而不是等到所有区域都可用才发布。

选择区域时的第四个考虑因素是成本。服务费用因使用服务的区域不同而异。例如，在 us-east-1 区域中运行 Amazon EC2 实例的费用可能与在 eu-west-1 区域中运行的费用不同。通常，成本差异可能不足以取代其他三个考虑因素。然而，如果区域之间在延迟、合规性或服务可用性方面的差异很小，则可以通过为您的环境使用成本较低的区域来节省资金。

最后，如果您的客户位于全球的不同地区，可以考虑将您的环境复制到多个离他们较近的区域，以优化客户体验。因为，这样您就可以将负载分配到多个环境中，每个环境中的组件成本就会降低，即使您添加更多基础设施也没问题。例如，添加第二个应用程序环境可能会让您在每个环境中的处理和存储容量要求都降低一半。AWS 旨在为您提供这种灵活性，而且您只需按实际用量付费，因此您可以将现有环境缩减，用节约下来的费用来添加另一个环境。

这种方法的缺点是，您现在需要管理两个环境。此外，并非所有组件都能缩减，并且缩减量足以弥补新组件的所有成本。此外，您可能需要在一个区域中维护一个单一存储“真实数据源”，例如主 Amazon Relational Database Service (Amazon RDS) 实例。您的辅助区域需要与存储实例进行通信，这可能会增加这些操作的延迟和成本。

模块 3 – 挑战实验：为咖啡馆创建静态网站

您现在将完成模块 3 – 挑战实验：为咖啡馆创建静态网站。

业务需求：一个简单的网站

Sofía 曾向 Nikhil 提到过，她想要的网站能够以图片展示咖啡馆的氛围，并为顾客提供详细的业务信息。

Frank 喜欢网站的想法。他一直在拍照，可以用来突出咖啡馆的菜单项目。

Sofía 向 Nikhil 提到过，她想要的网站能够以图片展示咖啡馆的氛围。该网站还应向客户提供业务详细信息，例如商店的位置、营业时间和电话号码。

Nikhil 很高兴为咖啡馆创建第一个网站。在本活动中，您将扮演 Nikhil，努力打造出咖啡馆的每个人都期待您能带来的结果。也许您还能超出他们的预期！

挑战实验：任务

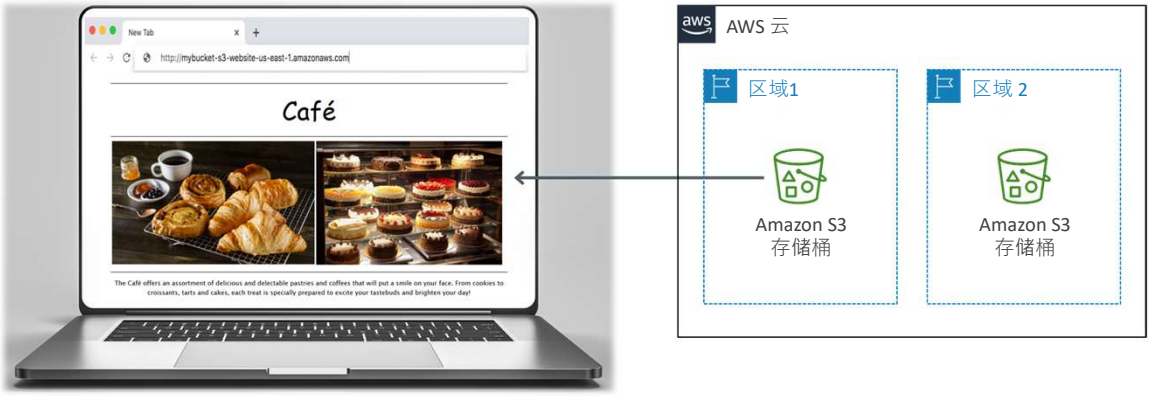
1. 提取此实验所需的文件
2. 创建 S3 存储桶来托管静态网站
3. 将内容上传到 S3 存储桶
4. 创建存储桶策略以授予公共读取访问权限
5. 为 S3 存储桶启用版本控制
6. 设置生命周期策略
7. 启用跨区域复制

在本挑战实验中，您将完成以下任务：

1. 提取此实验所需的文件
2. 创建 S3 存储桶来托管静态网站
3. 将内容上传到 S3 存储桶
4. 创建存储桶策略以授予公共读取访问权限
5. 为 S3 存储桶启用版本控制
6. 设置生命周期策略
7. 启用跨区域复制

挑战实验：最终产品

http://<bucket-name>.s3-website-<region>.amazonaws.com



在本挑战实验中，您将为咖啡馆创建一个静态网站。该网站将托管在 Amazon S3 上。创建 S3 存储桶并正确配置用于托管网站后，Web 浏览器应该能够使用指定的 Amazon S3 终端节点 URL 直接访问网站。

内容说明：架构图显示了两个区域，每个区域都有一个 S3 存储桶。一个 S3 存储桶指向咖啡馆的网站。http://<bucket-name>.s3-website-<region>.amazonaws.com。**内容说明结束。**

大约 60 分钟

现在可以开始挑战实验了。

挑战实验总结：要点

完成这个挑战实验之后，您的讲师现在可能会带您讨论此挑战实验的要点。

现在该复习本模块，并完成最后的知识考核和对实践认证考试问题的讨论了。

模块总结

总的来说，在本模块中，您学习了如何：

- 识别 Amazon Simple Storage Service (Amazon S3) 可以解决的问题
- 描述如何使用 Amazon S3 高效存储内容
- 了解各种 Amazon S3 存储类和成本注意事项
- 描述如何将数据移入和移出 Amazon S3
- 描述如何选择区域
- 创建静态网站

aws

总的来说，在本模块中，您学习了如何：
- 识别 Amazon Simple Storage Service (Amazon S3) 可以解决的问题
- 描述如何使用 Amazon S3 高效存储内容
- 了解各种 Amazon S3 存储类和成本注意事项
- 描述如何将数据移入和移出 Amazon S3
- 描述如何选择区域
- 创建静态网站

完成知识考核

现在该完成本模块的知识考核了。

考试样题

公司销售人员每天上传他们的销售数据。解决方案架构师需要一种用于这些文档的持久存储解决方案，同时还要防止用户意外删除重要文档。

哪种操作可以防止意外的用户操作？

选项	答案
A	将数据存储在 EBS 卷中并每周创建一次快照。
B	将数据存储在 S3 存储桶中并启用版本控制。
C	将数据存储在不同 AWS 区域的两个 S3 存储桶中。
D	将数据存储在 EC2 实例存储上。

思考答案选项，并根据关键词排除错误选项。

考试样题答案

公司销售人员每天上传他们的销售数据。解决方案架构师需要一种用于这些文档的持久存储解决方案，同时还要防止用户意外删除重要文档。

哪种操作可以防止意外的用户操作？

正确的答案是 B。

问题的关键词是持久存储解决方案、防止用户意外删除，以及哪种操作会提供保护。

以下是要识别的关键词：持久存储解决方案、防止用户意外删除，以及哪种操作会提供保护。

正确答案是 B。如果删除了某个受版本控制的对象，仍然可以通过检索最终版本来恢复该对象。

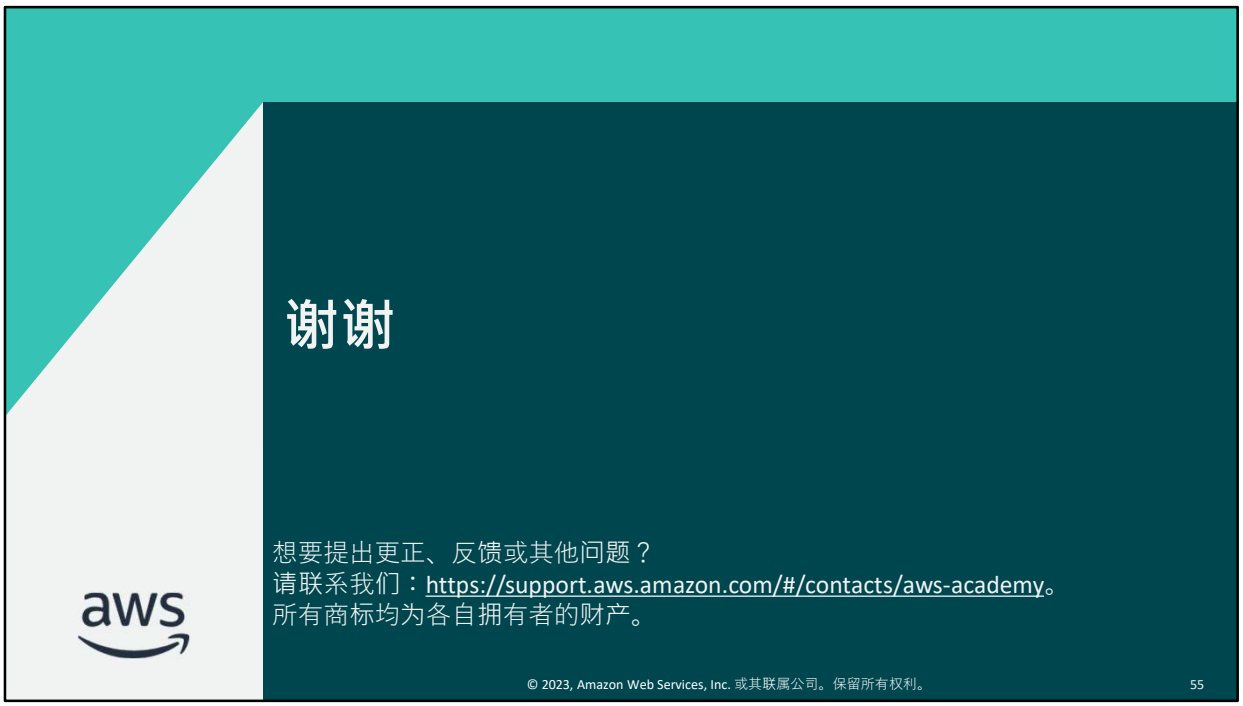
错误的答案：
选项 A 将丢失自上一个快照以来提交的任何更改。将数据存储在两个 S3 存储桶中（选项 C）提供的保护稍好一些，但用户仍然可以从两个存储桶中删除对象。EC2 实例存储（选项 D）是临时存储，永远不应用于需要持久性的数据。

其他资源

- Amazon S3 开发人员指南
- Amazon S3 常见问题
- Amazon S3 常见使用场景
- AWS 存储服务白皮书
- Amazon S3 存储类比较
- Amazon S3 阻止公开访问

如果您想进一步了解本模块中涵盖的主题，以下额外资源可能会对您有所帮助：

- Amazon S3 开发人员指南
- Amazon S3 常见问题
- Amazon S3 常见使用场景
- AWS 存储服务白皮书
- Amazon S3 存储类比较
- Amazon S3 阻止公开访问



感谢您完成本模块的学习。