企业备份与恢复

由内部设施到 AWS

Curd Zechmeister, Alex Tomic, Radhika Ravirala, Jeff Nunn

2014 年 12 月

目录

摘要

众多企业正在努力构建并部署一套综合性、基于云的备份与恢复策略，旨在对内部系统或者运行于托管环境下的系统加以保护。本份白皮书将提供此类混合型备份架构的设计与构建最佳实践，用于指导大家利用 Amazon Web Services（简称 AWS）存储及计算服务实现上述目标。本份白皮书专门用于指导企业业务相关备份与恢复、灾难恢复以及存储流程的管理工作。我们将在其中提供多项参考架构，用于指导大家顺畅完成设计流程，并将 AWS 云作为备份及恢复环境引入您的内部系统。

简介

对于大多数企业业务，内部备份与恢复解决方案往往需要占用大量资源，且其投资回报率非常有限。时至今日，众多企业开始利用云技术支撑备份与恢复解决方案，而不再需要自行构建并维护复杂且成本高昂的内部环境。

Amazon Web Services (简称 AWS)为备份与恢复需求提供多种广泛的高安全性、可扩展且极具成本效益的存储选项。大家可以利用各类 AWS 服务强化现有本地备份与恢复环境，亦可利用 AWS 服务单纯依靠云环境构建解决方案。在本份白皮书中，我们将探讨各可用选项以及如何帮助大家选择 AWS 服务以支持自有业务。

传统备份与恢复方法

首先，我们着眼于典型场景，即客户普遍反映的需要对自身备份与恢复环境加以简化，同时利用基于云类解决方案的最佳实践。大多数传统备份技术使用的是基于磁带的系统解决方案，即利用传统备份技术基于磁盘或者虚拟磁带将数据快照写入其中，从而实现备份与恢复效果。

磁带系统

在典型的磁带备份架构当中，数据持久存在于网络连接存储或者本地磁盘存储介质当中。在设置的规划当中，收集到的数据将被写入备份服务器，而后传输至磁带介质，由其构建起规模庞大的站内磁带库。而在进行检索时，则由某种形式的机器人自动管理各份磁带或者由工作人员进行查找。大多数企业还会将关键性业务应用程序数据通过广域网发送至小型异地磁带库处。

尽管面向磁带介质的备份写入操作易于实现且复制简单，但其拥有一系列固有弊端，具体包括连续读取速度缓慢、无法实现频繁测试、备份时间窗口过长以及备份与恢复过程可能引发卡带故障等。

这些缺陷使得基于磁带的备份解决方案在应对灾难事故时表现不佳。如果大家目前正在内部环境中使用基于磁带的备份解决方案，那么将这部分负载迁移至云端不仅能够消除物理介质处理带来的复杂性因素，同时亦可显著降低物理存储磁性介质带来的高昂使用成本。

基于磁盘的备份

在磁盘备份（简称 B2D）方案当中，数据以一级存储设备的快照形式进行存储，或者首先以磁带形式写入至磁盘，而后再迁移至磁带进行长期存储。使用基于磁盘的备份方案能够带来高于磁带的速度表现，且可靠性与灵活性亦更为出色。

尽管 B2D 解决方案拥有众多优势，但使用磁盘替代磁盘会带来更为可观的使用成本。利用基于磁盘的备份解决方案，大家必须频繁以分卷的快照形式进行数据备份。因此，大家需要提升数据存储所使用的磁盘容量，这可能导致解决方案的成本超出用户的可承受范围。

虚拟磁带库(简称 VTL)

虚拟磁带库（简称 VTL）在本质上属于一套基于磁盘的文件存储，其能够模拟一套传统磁带介质。虚拟磁带库技术目前被频繁应用于内部环境当中以作为备份方案，在 VTL 的帮助下，企业客户能够利用本地或者远程站点将数据备份至备份服务器，并在这里将其长久保存在磁盘内并作为虚拟磁带卡加以管理。一般来讲，VTL 解决方案允许大家对备份数据进行优化以削减存储容量需求。当数据经由网络被发送至 VTL 处，或者在某些情况下预先进行数据移动，则待存储数据会经过重复数据删除与压缩处理。这些技术能够大大减少存储备份所需要的磁盘空间，或者降低备份数据的规模，从而更为轻松地将其经由广域网实现指向数据中心的传输。

用于备份与恢复的各项 Amazon 服务

AWS 提供一系列服务，可用于面向混合与纯托管解决方案构建备份与恢复架构。AWS 之上的存储服务包括多种块存储形式，具体有 Amazon Elastic Block Store（即 Amazon 弹性块存储，简称 Amazon EBS）、Amazon Simple Storage Service（简称 Amazon 简单存储服务，简称 Amazon S3）以及 Amazon Glacier。为了能够与内部解决方案进行紧密结合，AWS 还提供一项存储网关服务，名为 AWS Storage Gateway，其可作为内部装置或者在 Amazon Elastic Compute Cloud（即 Amazon 弹性计算云，简称 Amazon EC2）上发挥作用。另外，AWS Marketplace 中也包含多种第三方解决方案。为了提升数据在内部环境与云环境之间的迁移效率，AWS 提供多种选项将用户数据中心同 AWS 云间进行对接，具体包括 AWS VPN CloudHub 以及 AWS Direct Connect。以下章节将概括这些服务在备份与恢复领域的各自优势。

Amazon Elastic Compute Cloud (简称 Amazon EC2)

Amazon Elastic Compute Cloud (简称 Amazon EC2)属于一项 Web 服务，能够在云环境下提供可随意调整规模的计算容量。其设计目标在于面向开发人员与系统管理员简化 Web 规模计算。Amazon EC2 允许大家只为自己实际使用的资源量支付费用，从而变更计算资源的经济模式。Amazon EC2 还为开发人员及系统管理员提供各类工具，用于构建具备故障弹性的应用程序并将其与常见故障场景隔离开来。欲了解更多信息，请参阅 Amazon EC2 页面。

Amazon Simple Storage Service (简称 Amazon S3)

Amazon Simple Storage Service (简称 Amazon S3) 提供高度安全且极具可扩展性的对象存储服务。

大家可以利用 Amazon S3 控制台随时立足于任意网络位置存储并检索任意规模的数据。大家可以单独使用 Amazon S3，或者将其与 Amazon EC2、Amazon Elastic Block Storage（简称 Amazon EBS）、Amazon Glacier 以及第三方存储库及网关相配合，面向广泛的各类用例提供极具成本效益的对象存储支持。Amazon EBS 允许大家创建存储分卷，并将其附加至 Amazon EC2 实例当中。Amazon S3 特别适合用于数据备份，因为其具备极高的设计持久性（持久性可达 99.999999999%）且成本低廉。

Amazon S3 将数据作为对象存储在名为“存储桶”的资源当中。AWS Storage Gateway 及多种第三方备份解决方案允许大家随意管理 Amazon S3 对象。大家也可以根据需要在单一存储桶内保存任意数量的对象，同时在存储桶内写入、读取及删除各类对象。单一对象的最大体积为 5 TB。欲了解更多信息，请参阅 Amazon S3 页面。

Amazon Glacier

Amazon Glacier 是一项成本极低的云归档存储服务，能够为数据归档与在线备份场景提供安全且持久性突出的存储支持。为了保证低廉的使用成本，Amazon Glacier 对需要频繁访问的数据做出优化，大家能够在数小时之内实现数据检索。利用 Amazon Glacier，大家能够以可靠方式存储任意规模的数据，其每 GB 每月使用成本可低至 0.01 美元，这一水平远低于内部解决方案的存储成本。Amazon Glacier 适用于需要长期保留的备份数据存储，或者需要长期甚至无限期存储的归档数据。

由于大家只需要为自己实际使用的资源量付费，因此 Amazon S3 与 Amazon Glacier 能够帮助各位摆脱备份存储容量规划的需要。欲了解更多信息，请参阅 Amazon Glacier 页面。

Amazon Elastic Block Store (简称 Amazon EBS)

Amazon EBS 提供持久性块级存储分卷，可配合 Amazon EC2 实例在 AWS 云内共同使用。每套 Amazon EBS 分卷可自动在其可用区内进行复制，从而保护用户免受组件故障的影响，最终实现极高可用性及持久性。Amazon EBS 分卷亦为备份数据的存储与检索提供一致性及低延迟性能承诺。利用 Amazon EBS，大家可以在数分钟内完成使用量的规模伸缩调整，且只需要为实际用量支付低廉费用。存储网关可频繁使用 Amazon EBS 分卷，从而将备份数据持久存储于 AWS 云当中。欲了解更多详细信息，请参阅 Amazon EBS 页面。

AWS Storage Gateway

AWS Storage Gateway 能够将内部软件装置同基于云的存储机制相对接，从而在内部 IT 环境与 AWS 存储基础设施之间建立无缝化且高安全性的集成效果。欲了解更多信息，请参阅 AWS Storage Gateway 页面。

AWS Direct Connect

AWS Direct Connect 能够轻松建立连接内部环境与 AWS 之间的专用网络连接。利用 AWS Direct Connect，大家可以在 AWS 与您的内部数据中心、办公环境或者异地环境之间建立专有连接，从而有效降低网络成本、提升传输带宽并提供较互联网连接更具一致性的使用体验。欲了解更多信息，请参阅 AWS Direct Connect 页面。

AWS Import/Export

AWS Import/Export 能够利用便携式存储设备进行数据传输，从而加快大规模数据与 AWWS 云之间的移动速度。AWS Import/Export 可利用 AWS 高速内部网络将数据直接传入及传出存储设备，从而回避互联网造成的传输效率影响。对于大规模数据集，AWS Import/Export 的速度通常高于互联网传输且较升级现有网络连接更具成本效益。欲了解更多信息，请参阅 AWS Import/Export 页面。

常规混合型备份与恢复方案

要将您的内部数据备份至 AWS 云，大家可以从以下两种常规方案当中做出选择：

- 通过指向 AWS 平台的 API 调用将备份数据直接写入至 Amazon S3，而后通过直接经由互联网的安全 HTTP PUT 与 GET 请求完成备份数据的放置与检索。在这里，端点本身能够直接接入 Amazon S3 以写入并检索数据。
- 将备份数据写入至作为存储网关的中间设备，而后再由存储网关设备将数据移动至 AWS 云。欲了解更多与存储网关相关的信息，请参阅本份白皮书内的“存储网关角色”章节。网关技术相关解决方案通常采用混合架构机制，其中部分组件立足于内部环境，另一部分则来自 AWS 生态系统。

Storage Gateways 的角色定位

大家实际采取的云环境备份数据写入以及云端备份数据检索方式会给性能及使用成本带来巨大影响。举例来说，使用存储网关技术需要内部硬件的配合，而直接指向 Amazon S3 的数据写入机制则会占用大量互联网传输带宽。那么存储网关技术如何帮助客户创建更为简便易行的内部备份与恢复使用体验？

存储网关负责将内部环境与云端相对接。网关属于一类硬件或者软件装置，其能够充当企业内物理位置与远程 AWS 云端云存储间的中介机制。大家可以利用自有 IT 基础设施部署单一存储网关，或者在多个位置部署多套网关。

为了降低各业务位置之间的网络连接传输能力，最好能够提前对数据进行压缩与重复数据删除处理，从而更为轻松地通过广域网进行本地环境数据移动。备份产品与存储网关装置通常能够执行数据压缩，且在多数情况下重复数据删除为内置功能。压缩机制属于数据编码方案，相较于未压缩数据，其能够有效降低数据体积。重复数据删除则能够通过去除重复数据对数据加以进一步压缩。应用重复数据删除及数据压缩功能能够大幅降低重复数据的出现机率，同时显著减少存储空间需求。存储网关能够很好地完成这些任务。

大家可以利用 AWS Storage Gateway 以执行以上提到的各项任务。另外，大家也可以充分发挥各类存储网关技术优势（包括硬件与软件装置），包括在 AWS Marketplace 当中提供的 Amazon 合作伙伴网络（简称 APN）认证方案。举例来说，NetApp 与 CTERA 双方皆具备网关技术，能够提供可行的本地环境与云端连接方式。

在本份白皮书的剩余部分内，我们将专注于 AWS Storage Gateway，探讨如何利用网关技术设计混合型存储架构。

使用 AWS Storage Gateway 能够带来哪些助益？

- **高安全性—** AWS Storage Gateway 能够在通过 SSL 进行数据上传及下载时，对其进行加密。在 Amazon S3 当中，则利用 AES 256 进行闲置数据加密。
- **利用 Amazon S3 进行持久性支持 —** AWS Storage Gateway 能够在 Amazon S3 当中将数据作为 EBS 快照加以存储。Amazon S3 的设计方案能够在两套设施之内同时保有数据，从而以冗余方式实现多个服务区内多座数据中心间的数据同步存储。
- **兼容性 —** AWS Storage Gateway 能够提供业界标准的分卷格式（例如 iSCSI）或者虚拟磁带库（简称 VTL）。iSCSI 接口意味着我们不再需要对内部应用的架构进行重新设计。
- **成本效益—** 大家只需要为实际使用的资源付费。
- **与 Amazon EC2 相集成 —** AWS Storage Gateway 允许大家轻松将数据由内部应用以镜像形式指向运行在 Amazon EC2 之上的应用处。
- **网络效率—** AWS Storage Gateway 只对经过修改的数据进行上传，而且会对全部上传及下载数据进行压缩。

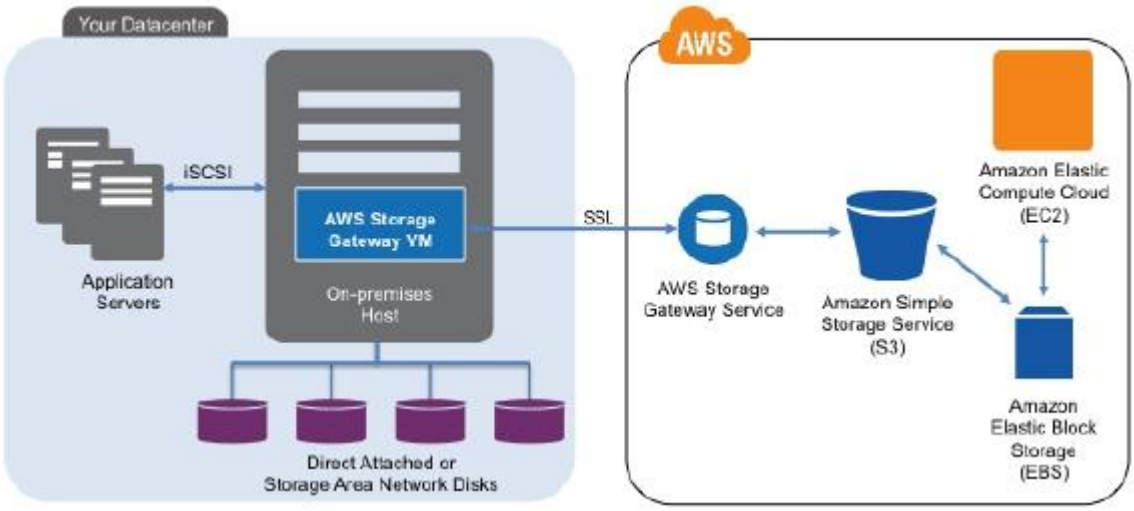
AWS Storage Gateway 配置选项

AWS Storage Gateway 支持以下三种配置：

- **网关缓存分卷—** 大家可以将自己的一级数据存储在 Amazon S3 当中，并将需要频繁访问的数据进行本地保留。网关缓存型分卷能够提供成本更为低廉的一级存储服务，从而最大程度降低内部存储规模要求，同时继续实现频繁访问数据的低延迟访问能力。
- **网关存储型分卷—** 如果大家需要对整体数据集进行低延迟访问，则可配置内部数据网关以本地存储一级数据，同时以异步形式将数据时间点快照备份至 Amazon S3。
- **网关虚拟磁带库(网关 VTL) —** 利用网关 VTL，大家可以以无限方式使用虚拟磁带。每套虚拟磁带可存储在由 Amazon S3 支持或者 Amazon Glacier 虚拟磁带架提供的虚拟磁带库当中。

AWS Storage Gateway 部署

以下示意图展示了 AWS Storage Gateway 在网关存储型分卷模式下采用的典型部署方式。



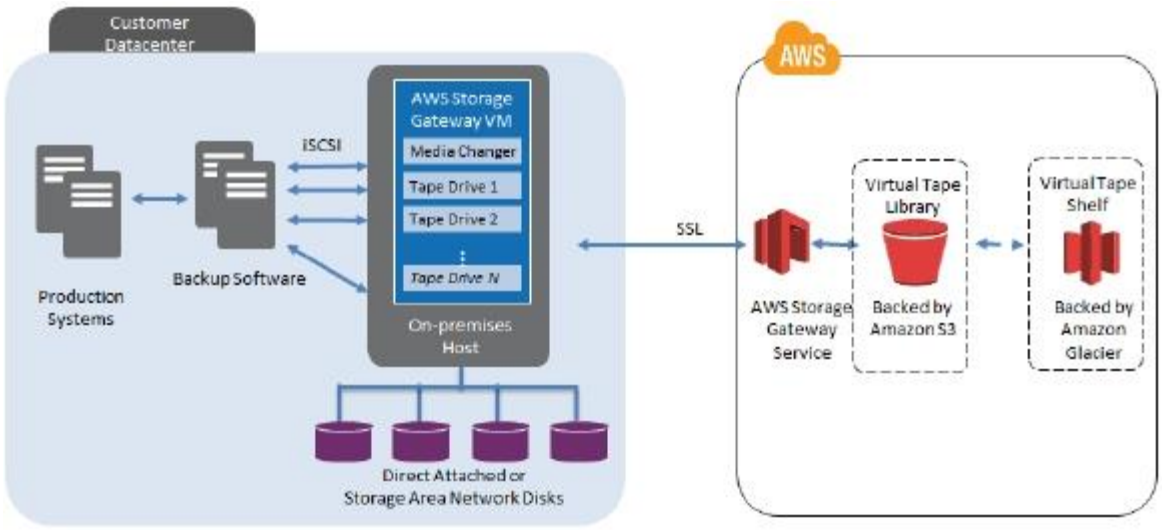
图一：

AWS Storage Gateway 的网关分卷模式部署示意图

以上示意图所示为部署在内部环境且作为虚拟装置运行在本地主机系统之上的 AWS Storage Gateway。在此种场景下，大家可以决定需要为 AWS Storage Gateway 附加多少接入存储资源，具体取决于网关实际使用方式。面向 AWS Storage Gateway 的通信通过安全嵌套层（简称 SSL）连接实现。数据被存储在 Amazon S3 当中，且能够保留在客户数据中心的系统当中。数据亦可保留在 Amazon EBS 设备内，其随后则可被附加至 Amazon EC2 实例处，例如借此实现灾难恢复策略。

AWS Storage Gateway 及其它存储网关技术也能够作为虚拟磁带库使用。利用网关 VTL，大家可以使用无数套虚拟磁带。每套虚拟磁带可存储在虚拟磁带库当中，而后者则由 Amazon S3 支持或者由 Amazon Glacier 实现的虚拟磁带架负责实现。该虚拟磁带库采用行业标准 iSCSI 接口，这意味着大家的备份应用能够在线访问各虚拟磁带。如果大家不再需要对容纳于虚拟磁带内的数据进行即时或者频繁访问，则可使用备份应用将其由虚拟磁带库迁移至虚拟磁带架，从而进一步降低存储成本。这套方案的一大关键性优势，在于客户能够继续使用已经部署于 IT 基础设施之内的备份与恢复软件应用，例如 Veeam、Backup Exec、TSM 以及 Robocopy，其可直接与行业标准及 iSCSI 兼容型虚拟磁带库进行通信。

以下示意图显示了一套典型网关拓扑，其部署于网关 VTL 模式之下。



图二：

作为行业标准 iSCSI VTL 进行部署的 AWS Storage Gateway

这套拓扑能够帮助大家将备份持续存储在内部环境下，从而利用 Amazon S3 实现低延迟访问以提供异地备份。如此一来，大家能够避免处理异地磁带存储或者 VTL 站到站复制带来的复杂性因素。

大家能够从 AWS 处将 AWS Storage Gateway 作为虚拟机进行下载，而后将其安装在选定的虚拟机管理程序当中，具体包括：VMware ESXi 或者微软 Hyper-V。如果大家正在部署内部网关，则可下载并部署该网关虚拟机，而后激活此网关。如果大家计划在 Amazon EC2 实例之上部署网关，则可启动一套包含有该网关虚拟机镜像的 Amazon Machine Image（简称 AMI），而后激活该网关。

作为最佳实践，大家应当考虑以本地方式附加至内部环境中存储网关的存储容量。以下表格汇总为 AWS Storage Gateway 本地存储容量规划时需要考虑的几项要点。

网关存储模式	考虑因素
网关存储型分卷	在网关存储分卷模式之下，完整分卷内的各快照存储于 AWS 云当中。因此，大家需要将必要磁盘空间的实际容量附加至该存储分卷。举例来说，如果大家计划从存储网关处为本地系统提供 50 TB iSCSI 分卷，则需要将 50 TB 本地存储附加至该网关。在这种模式下，您数据的一级副本以本地方式存储在 AWS Storage Gateway 之内，而各网关分卷的备份被存储在 AWS 之上。
网关缓存型分卷	网关缓存型分卷模式能够显著降低需要附加至存储网关的本地存储容量。在缓存型分卷模式当中，数据的一级副本被存储在 AWS 云之上，

网关存储模式	考虑因素

且只有缓存数据被存储于本地附加驱动器当中。存储网关上的所需存储容量取决于需要本地缓存的数据总量。

备份与恢复架构最佳实践

以下三套架构蓝图描述了对应的常见备份与恢复问题。从架构角度来看，完整的备份与恢复解决方案通常包含一款安装于端点之上的备份软件组件，同时配合一套内部网关解决方案。

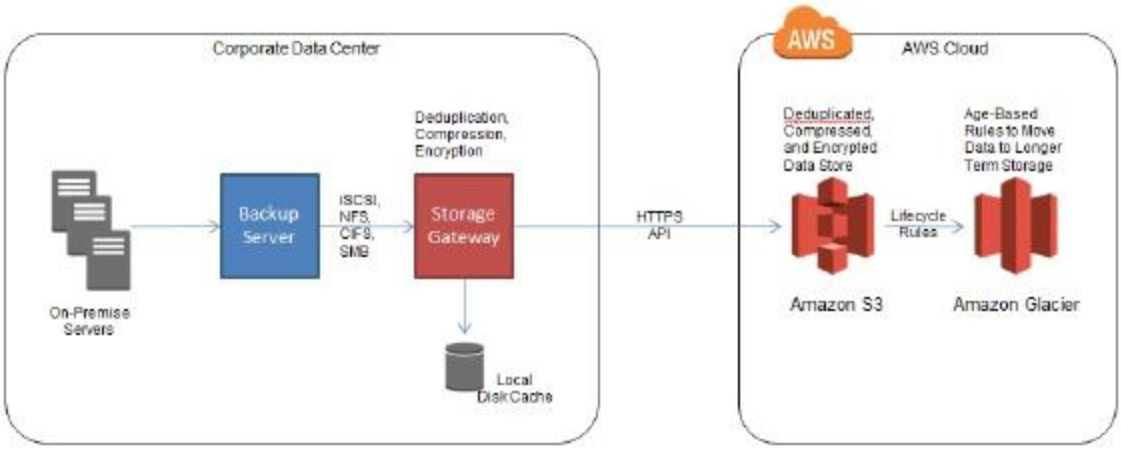
架构蓝图一：利用第三方网关实现内部备份

AWS Storage Gateway 以及来自 APN 合作伙伴的类似产品能够提供存储网关系统，其将内部备份基础设施同 AWS 当中的 Amazon S3 与 Amazon Glacier 服务加以结合。从概念层面讲，这些网关能够作为本地缓存利用 AWS 中的无限存储资源。在备份与恢复使用场景下，存储网关部署的主要目标在于取代磁带及磁盘存储系统等传统备份系统中的实现手段。通过这种方式，其将降低备份基础设施的复杂性并消除容量限制。

设计模式

此套架构蓝图能够将存储网关系统在主要数据中心内利用单一部署方式加以实现。企业备份服务器利用存储网关作为备份归档数据的一级存储。该存储网关能够提供一套本地磁盘缓存，从而改进备份性能并以异步方式将全部备份归档上传至 Amazon S3 当中。根据实际精选奈，备份数据亦可定期由 Amazon S3 处移动至 Amazon Glacier 以进行长期存储。

正如之前所提到，存储网关可以物理或者虚拟装置方式实现。要与备份基础设施加以集成，AWS Storage Gateway 等存储网关能够开放接口并与各类备份软件相兼容，具体包括 iSCSI 目标、CIFS/SMB 或者 NFS 共享、或者 VTL 模拟。各网关接口亦可配合 Amazon S3 或者 Amazon Glacier API 经由受 SSL 保护的 HTTPS 连接实现。



图三：存储网关与 AWS 相集成

性能注意事项

大家应当在评估存储网关解决方案时考虑以下两大性能因素：

- 备份服务器与存储网关之间的吞吐能力与数据传输带宽。
- 存储网关与 Amazon S3 之间的互联网带宽数据传输速率。

尽管 RTO 会受到这两项因素的影响，但 RPO 则主要受到接往 Amazon S3 的可用传输带宽的数据速率的影响。

备份服务器到存储网关

以下表格列出了影响备份服务器到存储网关间数据吞吐能力的常见因素。

因素描述

本地磁盘大小

本地磁盘存储容量越大，缓存容量也越大，这意味着备份服务器与存储网关之间的数据存储或者获取延迟更低。

本地磁盘性能

存储网关中的磁盘速度越快，由备份服务器向存储网关的备份数据交付速度越快。

CPU 与内存

计算资源越强大，软件功能执行速度越快，具体包括重复数据删除、压缩、加密以及广域网优化等任务。

因素	描述
局域网速度	局域网设备与以太网接口的速度越快，备份服务器到存储网关之间的数据吞吐能力越强。

备份服务器与存储网关之间的数据吞吐能力决定了备份时间窗口的长度。而数据吞吐率主要受到存储网关硬件的影响，例如磁盘子系统的容量与速度以及局域网连接的线路速度。在虚拟装置当中，该吞吐量则为分配给虚拟机的专用资源量。备份服务器软件与存储网关软件的效率亦在其中扮演着重要角色。

由存储网关到 AWS

以下表格显示了对存储网关到 AWS S3 之间连接性能存在重要影响的各项因素：

因素	描述
可用传输带宽	存储网关经由公共互联网或者直宫线路利用 HTTPS API 调用将数据存储至 Amazon S3 当中。
压缩	压缩机制能够降低数据对本地存储与 Amazon S3 的容量需求，同时减少存储网关与 Amazon S3 之间的数据传输总量。
重复数据删除	重复数据删除能够提供更高水平数据压缩能力，但同时可能在存储网关上战胜更多计算与内存资源。
广域网优化	广域网利用其它技术以降低存储网关与 Amazon S3 之间的传输带宽需求。

存储网关与 Amazon S3 之间的性能水平为满足恢复点目标（简称 RPO）的首要决定条件。这一性能不仅受到互联网连接速度与质量的影响，同时也涉及存储网关系统当中可能包含的软件功能效率。此类软件功能包括重复数据删除、压缩以及广域网优化。根据需要备份的内容类型的不同，上述技术亦能够将需要传输至 Amazon S3 以及 Amazon Glacier 存储端的数据总量降低高达 90%。

示例场景

大家必须权限互联网传输带宽、压缩、重复数据删除以及广域网优化效率等诸多因素，从而评估整体备份系统设计的实际效果。举例来说，我们假定面对这样一套备份系统设计：在重复数据删除与压缩之后，每天数据传输量为 50 GB，而互联网传输带宽为单一 T-1 线路且要求 RPO 为 24 小时。这套解决方案将无法起效，要求客户将传输带宽至少升级至 6 Mbps 以保证每天能够将全部备份数据在 24 小时之内移动至 Amazon S3。

安全性注意事项

以下表格汇总了大多数存储网关当中最为重要的各项安全功能。部分 APN 合作伙伴能够提供其它额外功能。

功能描述

传输时加密

全部数据应当通过受 SSL 保护的传输机制传入及传出 Amazon S3。

闲置时加密

存储在存储网关、Amazon S3 以及 Amazon Glacier 当中的全部数据都应当利用强加密手段进行加密。

密钥存储

存储网关可能允许大家将加密密钥存储在彼此分离的物理位置。

密钥轮换

存储网关应当提供选项以定期更新各加密密钥。

存储网关通过公共互联网与 AWS 进行交互，同时将数据存储在 Amazon S3 以及 Amazon Glacier 当中。大部分网关支持闲置数据加密；一般来讲，数据会利用 AES-128 或者 AES-256 算法进行加密，而后再作为持久信息备份归档于设备以及 Amazon S3 或者 Amazon Glacier 当中。大家可以将加密密钥保存在独立物理位置，从而提供额外安全性保障。由 APN 合作伙伴提供的特定存储网关还包含多种先进密钥轮换功能。

持久性与可用性

能够将数据长期存储于 Amazon S3 以及 Amazon Glacier 当中，存储网关能够获得高持久性支持。这两项服务能够以冗余方式将对象存储在跨越多套 Amazon S3 服务区设施内的各台设备，且能够在一年周期之内提供高达 99.999999999%对象持久性保障。

为了提升部署于数据中心内基础设施的可用性，大家可以同时配置两套存储网关，同时配置多条冗余路径以连接本地数据中心与 AWS。

Amazon S3 还提供一套可用性 SLA（即服务水平协议）。欲了解更多与 Amazon Said

可用性相关的信息，请参阅 Amazon S3 SLA 页面。

恢复流程

在恢复流程当中，大家需要在灾难恢复（简称 DR）位置设置一套存储网关。另外，大家还需要利用合适的互联网连接以下载经过压缩或者重复数据删除处理的数据，进行数据本地解密并将数据公布至备份服务器。作为最佳实践要求，大家还应当定期对恢复规程加以测试。

除了使用物理灾难恢复站点，大家还可以利用 AWS 作为潜在灾难恢复站点，因为这时您的数据已经保存在 AWS 当中。出于这一需求，大家可以立足于直接接入 Amazon 虚拟专有云（简称 Amazon VPC）直接接入 AWS Storage Gateway。目前 AWS Marketplace 中由 APN 合作伙伴提供的多数存储网关都预置有 AMI，允许大家在 AWS 当中实现存储网关方案的快速轻松部署。

架构蓝图二：利用网关实现多站点备份与恢复

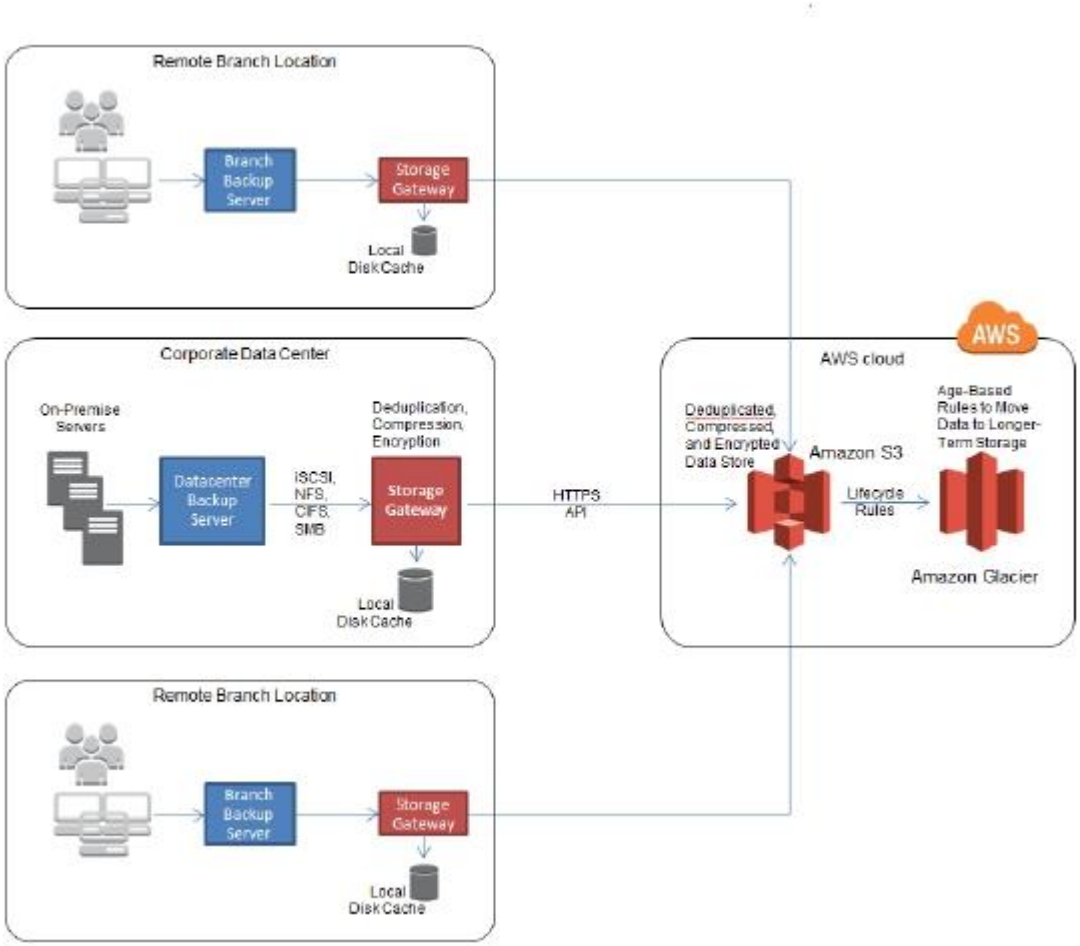
这套架构蓝图对上一蓝做出进一步扩展。其能够面向多套数据中心与远程分支环境实现备份需求，同时配合主业务数据中心内的备份基础设施。

设计模式

每套远程站点皆利用一台备份服务器与一套存储网关以将备份归档上传至 Amazon S3。AWS 可作为单一集中式备份方案，满足分布式企业业务中的备份需求。

存储网关可随时进行规模伸缩以匹配特定站点的需求。举例来说，由 10 名工作人员组成的远程销售办公环境所需要的设备规模自然远低于包含数百名员工的大型机构。其全局拓扑能够容纳多种规模不同的存储网关，从而实现最佳成本与性能平衡。

以下示意图展示了一套多站点架构，其中每套站点皆配备有存储网关以实现备份数据的快速访问。每套存储网关能够对来自 Amazon S3 的备份集进行存储与检索，其中数据将经过加密、重复数据删除以及解压缩。基于您所设置的具体控制政策，例如基于存在时长的生命周期策略，备份组可传入及传出 Amazon Glacier 以实现低成本长期保留。



图四：存储网关与 AWS 相集成

性能注意事项

远程站点，例如远程分支环境，通常互联网传输带宽较为有限，特别是与企业数据中心相比较。对于此类传输能力有限的场景，重复数据删除与压缩效率就成了实现 RPO 的关键。

备份时间窗口主要受到局域网速度及存储网关资源的影响。大家应当考虑面向不同远程站点设置必要的存储网关容量，从而保证备份时间窗口处于可接受范围之内。

恢复流程

恢复可立足于专有灾难恢复站点、AWS 之内或者其它远程站点实现。利用另一远程站点实现恢复的作法拥有诸多优势，这是因为其中已经包含各类必要的恢复实现条件：备份服务器、存储网关以及互联网连接。在这种情况下，大家可以立即对关键性文件及应用程序加以恢复。

持久性与可用性

相较于大型企业数据中心，远程站点往往在备份资源冗余与环境控制能力方面有所欠缺。利用 AWS 服务实现的备份机制能够提供备份数据与归档的持久性，即将数据同步至同一 AWS 服务区内多套设施之内的各台设备之上。

安全注意事项

物理安全在远程站点当中通常非常重要，特别是在小型远程分支环境当中。为了解决物理安全难题，大部分存储网关当中都包含有对应选项，旨在于数据被存储于本地磁盘缓存以及 Amazon S3 中之前对其进行加密。多数存储网关产品还包含其它功能，用于管理、轮换及保护加密密钥。

架构蓝图三：直接端点备份与恢复

The widespread use and adoption of endpoint devices—laptops, tablets, and smartphones—presents another challenge to IT: how to protect, backup from, and recover to devices within the remote workforce?各类端点设备的广泛应用——包括笔记本、平板电脑以及智能手机——给 IT 管理工作带来了新的挑战：如何

在常见场景当中，内部用户会利用自己的笔记本、台式机甚至是平板电脑或者智能手机创建各类内容。这些内容随后会以透明方式通过存储网关被备份到云环境当中，其间可进行加密、重复数据删除或者利用安全策略及访问控制列表加以保护。在存储完成之后，各文件可进行共享或者在企业当中供协作各方使用。

然而，远程用户通常不具备用于实现数据保护、备份、共享或者恢复的网关。以下选项能够帮助大家针对此类用户构建解决方案。

端点备份

异地用户及企业能够利用端点保护方案获得由存储网关实现的诸多收益，包括自动化与透明加密、重复数据删除、传输带宽限制以及计划同步等等。

各类高复杂度系统甚至允许我们搜索、下载以及恢复更多功能，同时为大型文件配置共享功能。

如以下示意图所示，大家的数据可利用单一软件解决方案交付至 AWS 云，且该方案能够根据需要轻松利用 APN 进行交互。

图五：利用 VPN 或者直连实现端点保护

被备份至 Amazon S3 的数据可以无缝化方式移动至 Amazon Glacier——这是一项成本极低的归档存储服务——并能够利用数据生命周期策略在特定时间段后完成数据归档。

端点恢复

端点设备的恢复用途多种多样，从简单的文件恢复到应用识别，从时间点恢复到完整的远程端点文件系统部署。本地存储的快照可定期进行规划，而端点恢复系统可用于以自动方式实现快照管理，从而立足于快照实现高效快速的恢复效果。我们建议大家将这些端点设备引入考量，从而规划自己的灾难恢复架构以及业务持续性方案。

在大多数情况下，远程用户会利用虚拟桌面基础设施（简称 VDI）以访问远程服务器上的桌面环境，且各状态文件——例如配置、偏好及其它本地存储——必须进行备份及恢复，从而实现最低恢复时间目标（简称 RTO）。

端点保护

除了数据备份与恢复，我们建议大家在存储与备份规划当中考虑到端点保护机制。端点保护能够利用多种数据丢失预防技术，从而避免未授权访问对数据的影响。此类技术具体包括：

- 端点设备可以远程方式由管理员进行内容清除。
- 管理员只能以远程方式对凭证或者敏感数据进行清除或者加密，而操作系统或者其它文件则不受影响。
- 地理位置服务可帮助追踪丢失或者被盗设备，从而保护数据泄露及企业信息。

通过在端点备份、恢复以及保护机制当中引入各项最佳实践，大家能够创建起一套既有效控制企业 IT 成本、又最大程度提升生产效率的数据保护策略。

基于云的备份模式

利用基于云的备份方案能够帮助大家免受磁带备份机制的种种限制，包括有限的存储容量与介质可靠性问题，但其同时也会在备份与恢复策略中引入新的元素。举例来说，相较于磁带库的物理安全性，云备份模式能够利用加密管理实现数据安全性。而相比于磁带库硬件可靠性，如今互联网传输带宽与可用性将成为实现备份性能与可靠性的核心途径。

注意事项

以下各项列举了 AWS 之上云备份与恢复解决方案中的设计注意事项，我们需要将其引入备份与恢复解决方案规划工作。

- **数据分类** — 您的企业是否具备数据分类策略，即区分出哪些数据类型应当被存储在云环境当中？举例来说，您是否会将需要采取特定加密级别的数据纳入备份组？或者是否拥有必须保存在特定地理位置的数据？考虑利用特定 AWS 服务区作为数据存储位置，考虑如何管理数据加密，同时根据数据分类结论决定哪些数据集合不应被存储在云环境之内。
- **加密密钥管理**— 那么 AWS 能否利用生成的密钥对闲置中的数据进行加密？您是否需要利用您自己的密钥进行数据加密？如果您需要管理自己的密钥，可以考虑使用 AWS CloudHSM 或者 AWS 密钥管理服务等密钥管理解决方案，也可以使用自己的现有内部密钥管理机制。另外，AWS Storage Gateway 当中亦内置有密钥管理，大家可利用其加密自己的备份数据。
- **网络传输带宽** — 确定满足备份与恢复要求所必需的网络连接类型。考虑直接经由互联网利用 HTTPS 进行端点备份，从而确保从位于世界任意位置的设备处获取备份，并恢复至任何设备处。对于内部环境，则可利用网关技术实现 AWS 备份，即考虑利用 AWS Direct Connect 的 VPN 连接。远程分支机构通常适合使用 VPN 连接，而中到大型数据中心则可从 AWS Direct Connect 处受益。
- **备份与恢复方法**— 是否需要对完整系统或者附加至各系统的分卷进行快照备份？举例来说，VMware 访客能够利用数据快照存储实现频繁存储。这些快照随后可经由 AWS Storage Gateway 分卷移动至 AWS 云当中。如果大家使用备份软件对系统进行文件级别备份，那么可以考虑使用网关上的缓存型分卷或者提供统一命名空间的解决方案，例如 CTERA 云服务交付平台。另外，大家也可以考虑使用传统备份与恢复方法进行数据复制。举例来说，将内部数据库复制至 AWS 之上的试点数据库，并将此作为主要备份与灾难恢复方法。欲了解更多与利用 AWS 实现灾难恢复的细节信息，请参阅利用 Amazon Web Services 实现灾难恢复白皮书。

· **磁带替换** — AWS 是一套强大的磁带替换方案。Amazon Storage Gateway 内置有虚拟磁带库（简称 VTL）功能，且内置于可随时部署的软件装置。如果大家需要在企业内部保留磁带概念，或者使用现有软件产品与 ISCSI 兼容 VTL 进行交互，则可考虑使用这项技术。AWS Storage Gateway 技术与 VTL 功能相配合能够帮助大家利用云资源取代磁带库。

· **备份数据集可用性**—考虑需要以怎样的速度访问特定备份数据集。举例来说，您是否需要在附加至存储网关的磁盘保留特定备份数据集，并利用本地级别网络速度实现即时恢复？如果答案是肯定的，那么大家可以考虑利用内部存储网关在缓存当中保留最新数据。如果大家可以等待数个小时以获取备份数据集，则应考虑直接在 Amazon Glacier 当中存储备份数据。

· **可访问性**— 将您的备份数据存储在 AWS 云当中意味着我们可以在任意位置利用互联网接入并实现访问。考虑最合适的数据访问位置，而后选择与潜在恢复位置距离最近的 AWS 服务区作为备份位置。另外，大家必须进一步考虑监管要求，从而选择数据在具体地理存储位置。

· **导入现有备份数据** — AWS 提供一种极具成本效益的方式将大量数据自物理介质处进行导入。AWS Import/Export 服务能够以 U 盘方式接收物理介质中的数据。AWS Import/Export 将创建一项清单文件，其负责描述接收及恢复数据所使用的具体介质。AWS Import/Export 服务允许大家以邮寄方式向 AWS 设施发送 U 盘以进行数据导入。另外，各位亦可以考虑首先将大规模现有备份数据写入至内部 U 盘当中，将其邮寄至 AWS，而后利用 AWS Import/Export 以快速经济的方式将数据加载至目标 Amazon S3 存储桶或者 Amazon EBS 分卷。

恢复：备份对复制

传统备份与恢复架构长久以来已经成为 IT 灾难恢复策略中的重要组成部分，且已经拥有一套成熟且完善的测试方法以确保数据安全。然而，随着网络传输带宽资源性价比的提升，企业可以重新考虑自己的备份与恢复策略，包括灾难恢复（简称 DR）策略。这一变化由原本的典型备份与恢复配置机制转移至经由备份与恢复实现的数据故障转移方向，特别是在灾难恢复场景之下。

在您的灾难恢复策略当中，我们建议大家采取以下方法：

- 传统备份与恢复
- 在 AWS 上建立试水项目
- 在 AWS 上实现热备份
- 在 AWS 上实现多站点

大家可以将上述几项模式加以结合，从而实现与企业要求相匹配的 RTO/RPO 目标。欲了解更多利用 AWS 实现灾难恢复的相关信息，请参阅利用 Amazon Web Services 实现灾难恢复白皮书。

总结

Amazon Web Services 能够为大家提供多种不同方案，用于实现基于云的备份及恢复。举例来说，大家可以将数据直接备份至 Amazon S3 当中，同时立足于内部或者云环境下端点进行数据恢复。作为另一种选项，大家也可以将内部备份及恢复解决方案全部替换为存储网关技术，例如 AWS Storage Gateway 及云备份软件。大家亦能够通过使用低成本 Amazon S3 与 Amazon Glacier 等对象存储长期保留备份数据集，从而显著降低备份解决方案的 TCO。

扩展阅读

欲获取更多帮助，请参考以下资料：

Amazon S3 上手指南：
http://docs.amazonwebservices.com/AmazonS3/latest/gsg/

Amazon EC2 上手指南：
http://docs.amazonwebservices.com/AWSEC2/latest/GettingStartedGuide/

AWS 合作伙伴目录 (包含一份 AWS 解决方案供应商列表):
http://aws.amazon.com/solutions/solution/providers/

AWS 安全性与合规性中心：
http://aws.amazon.com/security/

AWS 架构中心：
http://aws.amazon.com/architecture

利用 AWS 实现灾难恢复：
http://media.amazonwebservices.com/AWS_Disaster_Recovery.pdf

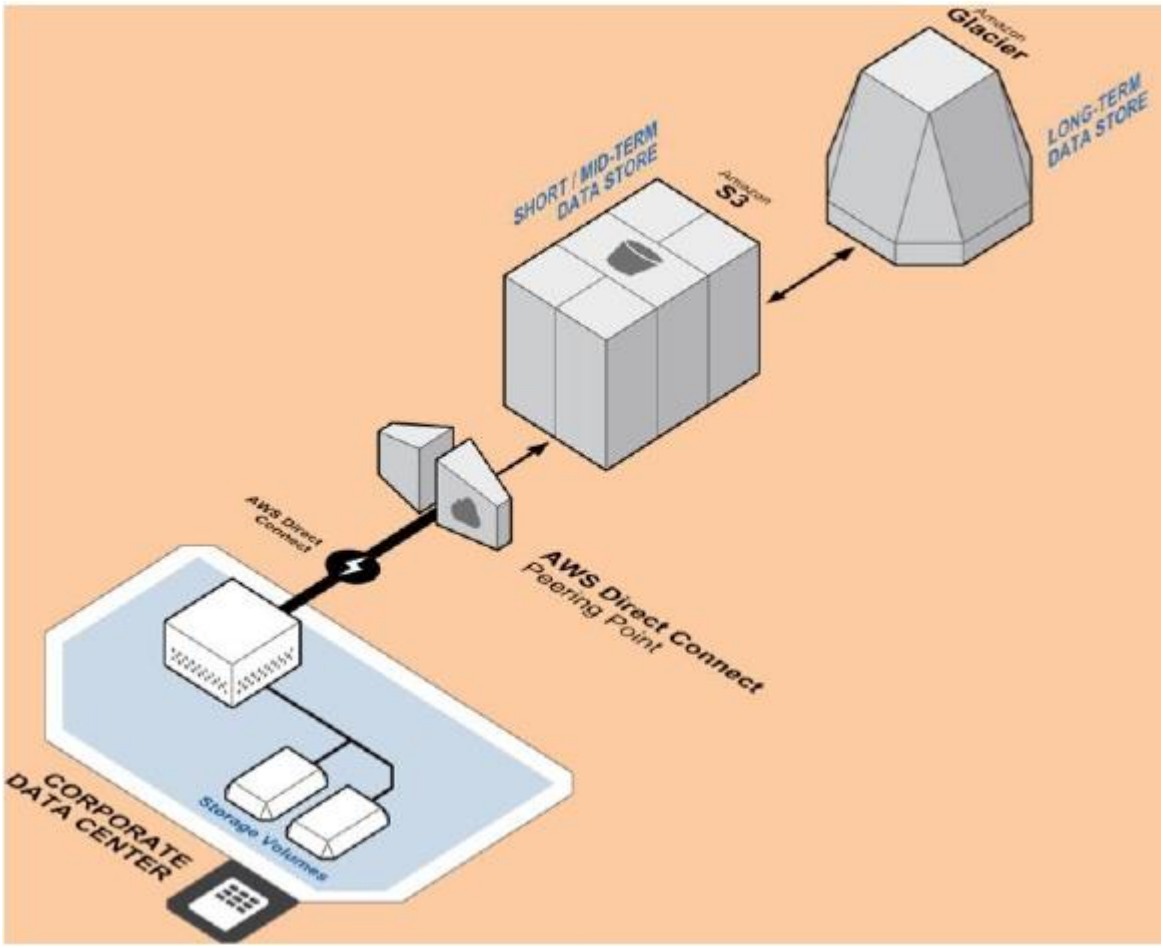
通告

附录：经由 AWS Direct Connect 由内部接入 AWS

备份与恢复流量本身规模很大，同时需要耗费大量时间。举例来说，特定业务解决方案的备份数据需要在可预测的时间段内完成，从而满足灾难恢复方案的 RPO 要求。另外，将备份数据存储在 AWS 云内的备份与恢复架构需要考虑网络条件限制，即数据经由互联网直接传输时占据的带宽资源。作为可行方案之一，大家可以利用 AWS Direct Connect 在内部系统与 AWS 之间进行网络控制。

大家可以利用 AWS Direct Connect 对内部环境与 AWS 云之间的网络集成进行严密控制。作为最佳实践之一，大家应当量化自己的基准数据传输需求，并确保其匹配内部备份与恢复工作负载的实际要求。Amazon 解决方案架构师（Amazon Solutions Architect）能够帮助大家找到理想的连接方法，并可帮助各位掌握广域网连接所必需的网络传输带宽。

AWS Direct Connect 能够利用标准 1 GBit 或者 10 GBit 以太网光纤线缆将内部网络与 AWS Direct Connect 位置进行对接。该线缆的一端接入您的路由器，另一端则接入 AWS Direct Connect 路由器。利用这套连接机制，大家能够直接面向 Amazon VPC 创建虚拟接口，从而绕过网络路径中的互联网服务供应商。对于时间敏感型备份与大规模数据，AWS Direct Connect 允许大家通过专用容量连接实现可预测的云备份与恢复解决方案。以下示意图所示为一套典型的 AWS Direct Connect 拓扑结构。



图六：利用 AWS Direct Connect 实现备份与恢复

以上示意图显示一套利用专有 AWS Direct Connect 对对等点接入 APN 合作伙伴位置的内部数据中心。在对等点处，大家可以利用 1 GBit 或者 10 GBit 连接接入 AWS 云，而后实现快速访问 Amazon S3 以及 Amazon Glacier。

自推出 AWS Direct Connect 以来，AWS 的对等点数量一直在不断增加。以下表格显示了当前由 APN 合作伙伴提供的对等位置。

AWS Direct Connect Location	AWS Region
CoreSite NY1 & NY2	US East (Virginia)

AWS Direct Connect Location	AWS Region
CoreSite One Wilshire & 900 North Alameda, CA	US West (Northern California)
Equinix DC1 – DC6 & DC10	US East (Virginia)
Equinix SV1 & SV5	US West (Northern California)
Equinix SE2 & SE3	US West (Oregon)
Eircom, Clonshaugh	EU West (Ireland)
TelcityGroup, London Docklands	EU West (Ireland)

我们建议大家利用 AWS Direct Connect 传输大规模备份与恢复数据集。然而为了控制实现成本，我们同样建议大家限定备份方案所使用的互联网传输带宽，确保其仅提供与备份窗口及 RTO/RPO 要求相符的资源容量。通过这种方式，大家将能够尽可能降低支付给互联网服务供应商（简称 ISP）的费用，同时避免互联网传输带宽提升或者新合约带来的额外成本。再有，所有经由 AWS Direct Connect 进行传输的数据应当尽可能加以规模缩减，而非直接经由互联网传输，旨在显著降低网络成本。

文档修订

2014 年 12 月 12 日 – 初始文本