NetApp 混合云数据解决方案—基于客户用例的 Spark 和 Hadoop

NetApp Solutions

NetApp
September 10, 2024

This PDF was generated from https://docs.netapp.com/zh-cn/netapp-solutions/data-analytics/hdcs-sh-solution-overview.html on September 10, 2024. Always check docs.netapp.com for the latest.

目录

NetApp 混合云数据解决方案—基于客户用例的 Spark 和 Hadoop . . . 1
TR-4657： NetApp 混合云数据解决方案—基于客户用例的 Spark 和 Hadoop . . . 1
NetApp 为大数据架构提供支持的 Data Fabric . . . 2
Hadoop 数据保护和 NetApp . . . 5
Hadoop 数据保护用例概述 . . . 5
用例 1： 备份 Hadoop 数据 . . . 6
用例 2： 从云到内部环境的备份和灾难恢复 . . . 10
用例 3： 对现有 Hadoop 数据启用 DevTest . . . 12
用例 4： 数据保护和多云连接 . . . 13
用例 5： 加快分析工作负载的速度 . . . 14
结论 . . . 16

NetApp 混合云数据解决方案—基于客户用例的 Spark 和 Hadoop

TR-4657： NetApp 混合云数据解决方案—基于客户用例的 Spark 和 Hadoop

NetApp 公司 Karthikeyan Nagalingam 和 Sathish Thyagarajan

本文档介绍使用 NetApp AFF 和 FAS 存储系统， NetApp Cloud Volumes ONTAP ，NetApp 互联存储以及适用于 Spark 和 Hadoop 的 NetApp FlexClone 技术的混合云数据解决方案。客户可以通过这些解决方案架构为其环境选择合适的数据保护解决方案。NetApp 在与客户及其业务用例进行交互的基础上设计了这些解决方案。本文档提供了以下详细信息：

- 为什么我们需要针对 Spark 和 Hadoop 环境以及客户面临的挑战提供数据保护。
- 由 NetApp 愿景提供支持的 Data Fabric 及其组件和服务。
- 如何使用这些组件构建灵活的数据保护工作流。
- 根据实际客户使用情形确定的多个架构的优缺点。每个用例都包含以下组件：
  - 客户情形
  - 要求和挑战
  - 解决方案
  - 解决方案摘要

为什么选择 Hadoop 数据保护?

在 Hadoop 和 Spark 环境中，必须解决以下问题：

- * 软件或人为故障。 * 执行 Hadoop 数据操作时，软件更新中出现人为错误，可能导致出现错误行为，发生原因从而可能导致作业产生意外结果。在这种情况下，我们需要保护数据，以避免出现故障或出现不合理的结果。例如，由于对流量信号分析应用程序进行的软件更新执行不当，这是一项新功能，无法以纯文本的形式正确分析流量信号数据。该软件仍会分析 JSON 和其他非文本文件格式，从而导致实时流量控制分析系统生成缺少数据点的预测结果。这种情况可能会导致发生原因输出出现故障，从而可能导致交通信号发生意外。通过提供快速回滚到先前工作应用程序版本的功能，数据保护可以解决此问题描述问题。
- * 大小和规模。 * 由于数据源和卷的数量不断增加，分析数据的大小每天都在增长。社交媒体，移动应用程序，数据分析和云计算平台是当前大数据市场中的主要数据源，该市场的增长速度非常快，因此需要对数据进行保护，以确保准确的数据运行。
- * Hadoop 的原生数据保护。 * Hadoop 具有一个原生命令来保护数据，但此命令不会在备份期间提供数据的一致性。它仅支持目录级备份。Hadoop 创建的快照为只读快照，不能用于直接重复使用备份数据。

Hadoop 和 Spark 客户面临的数据保护挑战

Hadoop 和 Spark 客户面临的一个常见挑战是，在数据保护期间，在不对生产集群性能产生负面影响的情况下，缩短备份时间并提高备份可靠性。

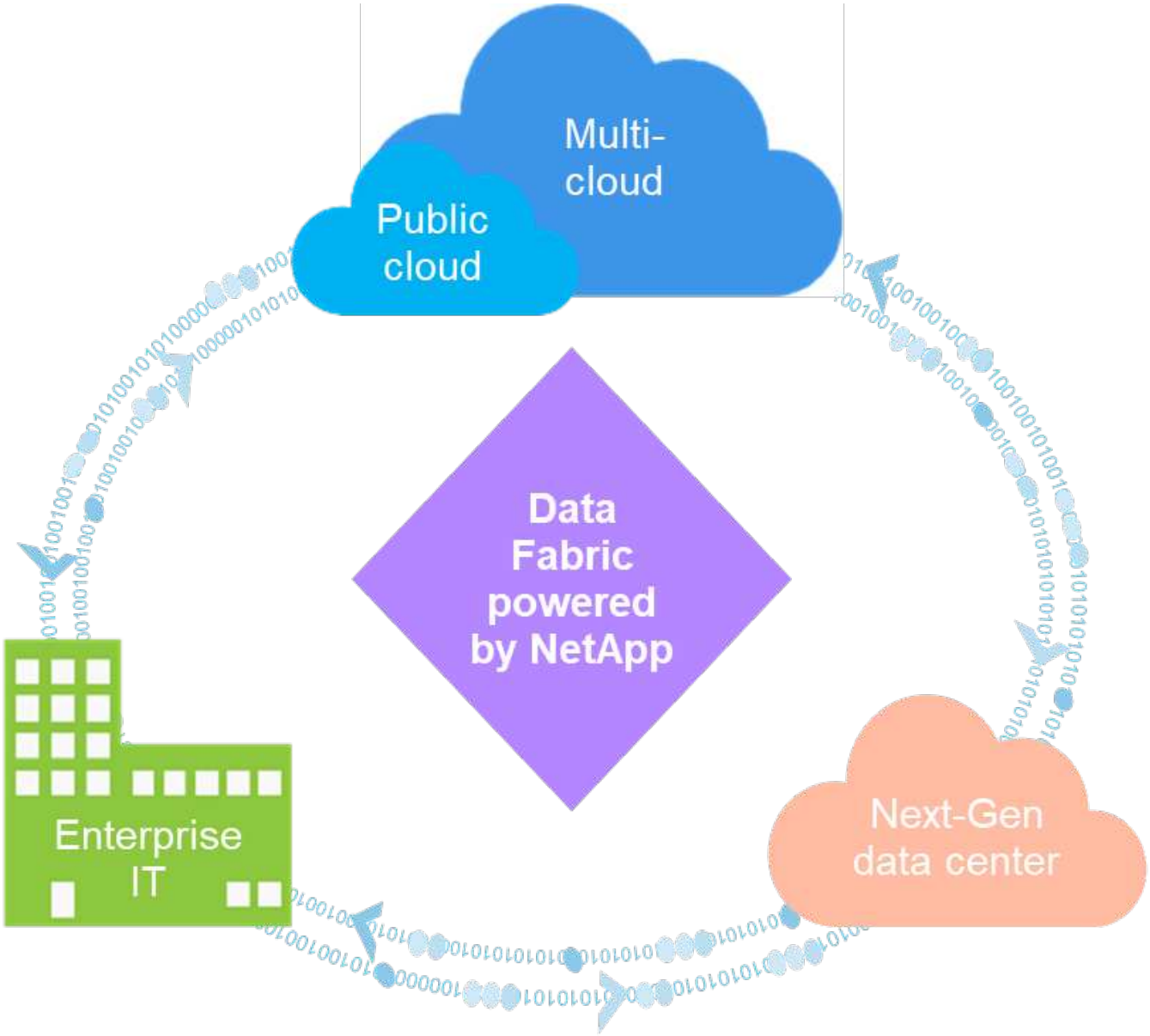
客户还需要最大限度地减少恢复点目标（RPO）和恢复时间目标（RTO）停机时间，并控制其内部和基于云的灾难恢复站点，以实现最佳业务连续性。这种控制通常来自企业级管理工具。

Hadoop 和 Spark 环境非常复杂，因为不仅数据量庞大且不断增长，而且数据的到达速度也在不断提高。在这种情况下，很难从源数据快速创建高效，最新的 DevTest 和 QA 环境。NetApp 认识到这些挑战，并提供了本白皮书中介绍的解决方案。

NetApp 为大数据架构提供支持的 Data Fabric

由 NetApp 提供支持的 Data Fabric 可简化并集成云环境和内部环境中的数据管理，以加速数字化转型。

由 NetApp 提供支持的 Data Fabric 可提供一致且集成的数据管理服务和应用程序（组件），以实现数据可见性和洞察力，数据访问和控制以及数据保护和安全，如下图所示。



经验证的 **Data Fabric** 客户用例

由 NetApp 提供支持的 Data Fabric 为客户提供了以下九个经验证的使用情形：

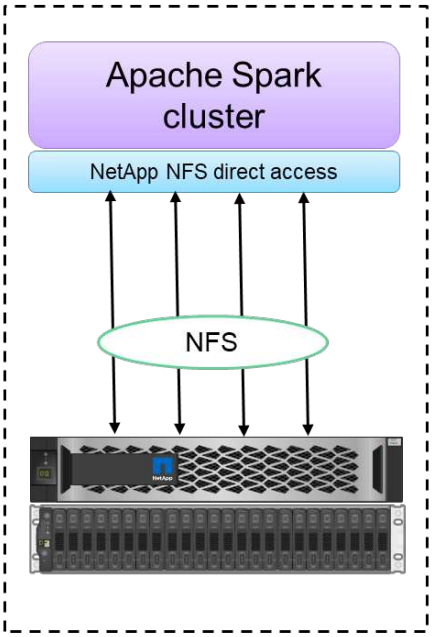
- 加快分析工作负载的速度
- 加快开发运营转型
- 构建云托管基础架构
- 集成云数据服务
- 保护和保护数据
- 优化非结构化数据
- 提高数据中心效率
- 提供数据洞察力和控制力
- 简化和自动化

本文档涵盖九个使用情形中的两个（及其解决方案）：

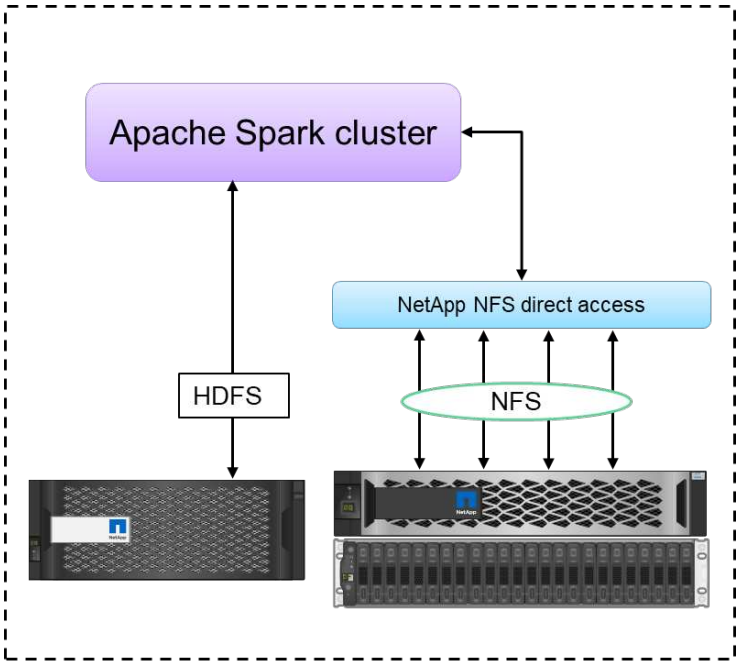
- 加快分析工作负载的速度
- 保护和保护数据

NetApp NFS 直接访问

借助NetApp NFS、客户可以对现有或新的NFSv3或NFSv4数据运行大数据分析作业、而无需移动或复制数据。它可以防止多个数据副本，并且无需将数据与源进行同步。例如，在金融领域，将数据从一个位置移动到另一个位置必须履行法律义务，这不是一项容易的任务。在这种情况下， NetApp NFS 直接访问会分析其原始位置的财务数据。另一个主要优势是，使用 NetApp NFS 直接访问可通过使用原生 Hadoop 命令简化对 Hadoop 数据的保护，并利用 NetApp 丰富的数据管理产品组合支持数据保护工作流。



Configuration 1: NFS as primary storage



Configuration 2: HDFS and NFS in single Spark cluster

NetApp NFS 直接访问为 Hadoop/Spark 集群提供了两种部署选项：

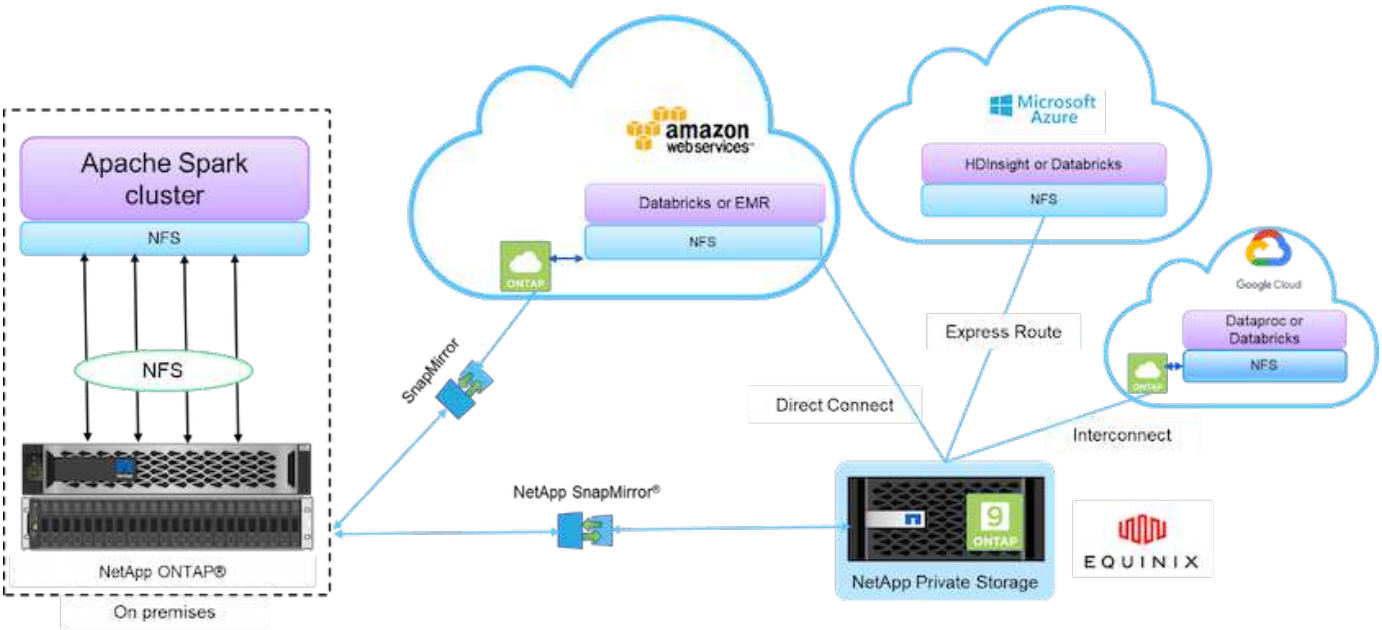
- 默认情况下， Hadoop/Spark 集群使用 Hadoop 分布式文件系统（ HDFS ）进行数据存储和默认文件系统。NetApp NFS 直接访问可以将默认 HDFS 替换为 NFS 存储作为默认文件系统，从而对 NFS 数据执行直接分析操作。
- 在另一种部署选项中， NetApp NFS 直接访问支持在一个 Hadoop/Spark 集群中将 NFS 与 HDFS 配置为额外的存储。在这种情况下，客户可以通过 NFS 导出共享数据，并从同一集群访问数据以及 HDFS 数据。

使用 NetApp NFS 直接访问的主要优势包括：

- 分析当前位置的数据，防止执行将分析数据移动到 HDFS 等 Hadoop 基础架构这一耗时且性能极高的任务。
- 将副本数量从三个减少为一个。
- 支持用户分离计算和存储，以便独立扩展。
- 利用 ONTAP 丰富的数据管理功能提供企业数据保护。
- 已通过 Hortonworks 数据平台认证。
- 支持混合数据分析部署。
- 利用动态多线程功能缩短备份时间。

大数据的组件

由 NetApp 提供支持的 Data Fabric 集成了数据管理服务和应用程序（组件），用于实现数据访问，控制，保护和安全性，如下图所示。



上图中的组件包括：

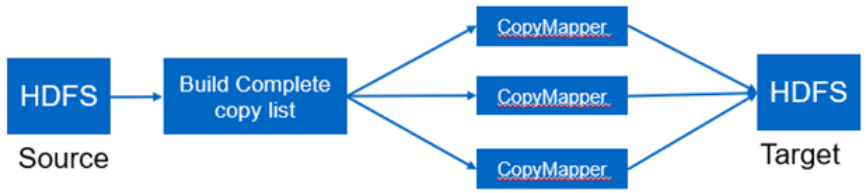
- * NetApp NFS 直接访问。 * 为最新的 Hadoop 和 Spark 集群提供对 NetApp NFS 卷的直接访问，而无需额外的软件或驱动程序要求。
- * NetApp Cloud Volumes ONTAP 和云卷服务。 * 基于在 Microsoft Azure 云服务的 Amazon Web Services （ AWS ）或 Azure NetApp Files （ ANF ）中运行的 ONTAP 的软件定义的连接存储。

- * NetApp SnapMirror 技术 * 。在内部部署和 ONTAP 云或 NPS 实例之间提供数据保护功能。
- * 云服务提供商。 * 这些提供商包括 AWS ， Microsoft Azure ， Google Cloud 和 IBM Cloud 。
- * PaaaS 。 * 基于云的分析服务，例如 AWS 中的 Amazon Elastic MapReduce （ EMR ）和 Databricks 以及 Microsoft Azure HDInsight 和 Azure Databricks 。

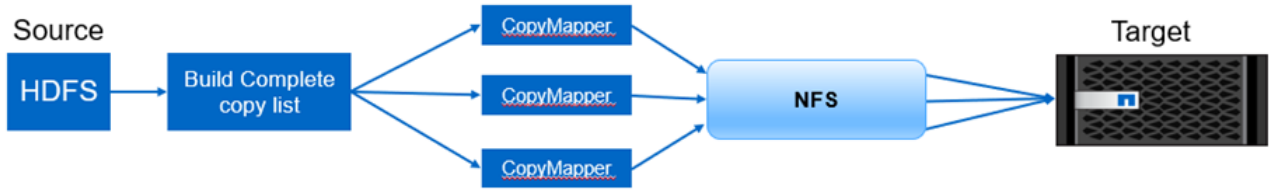
Hadoop 数据保护和 NetApp

Hadoop DistCp 是一款原生工具，用于大型集群间和集群内复制。下图所示的 Hadoop DistCp 基本过程是一个典型的备份工作流，它使用诸如 MapReduce 等 Hadoop 原生工具将 Hadoop 数据从 HDFS 源复制到相应目标。

通过 NetApp NFS 直接访问，客户可以将 NFS 设置为 Hadoop DistCp 工具的目标，以便通过 MapReduce 将数据从 HDFS 源复制到 NFS 共享。NetApp NFS 直接访问可用作 DistCp 工具的 NFS 驱动程序。



Hadoop Distcp Basic Process



Hadoop Distcp and NetApp

Hadoop 数据保护用例概述

本节将简要介绍数据保护使用情形的问题描述，这是本白皮书的重点内容。其余部分将详细介绍每个使用情形，例如客户问题（场景），要求和挑战以及解决方案。

用例 1： 备份 Hadoop 数据

在此使用情形中、NetApp NFS卷帮助一家大型金融机构将备份时间从超过24小时缩短到几小时以内。

用例 2： 从云到内部环境的备份和灾难恢复

一家大型广播公司利用 NetApp 提供支持的 Data Fabric 作为组件，根据不同的数据传输模式（例如按需，即时， 或基于 Hadoop/Spark 集群负载。

用例 3： 对现有 Hadoop 数据启用 DevTest

NetApp 解决方案帮助在线音乐分销商在不同的分支中快速构建多个节省空间的 Hadoop 集群，以便使用计划的策略创建报告并运行每日 DevTest 任务。

用例 4： 数据保护和多云连接

一家大型服务提供商使用由 NetApp 提供支持的 Data Fabric ，从不同的云实例为其客户提供多云分析。

用例 5： 加快分析工作负载的速度

最大的金融服务和投资银行之一使用 NetApp 网络连接存储解决方案来缩短 I/O 等待时间并加快其量化金融分析平台的运行速度。

用例 1： 备份 Hadoop 数据

在这种情况下，客户拥有一个大型内部 Hadoop 存储库，并希望对其进行备份以实现灾难恢复。但是，客户当前的备份解决方案成本高昂，并且备份时间过长，超过 24 小时。

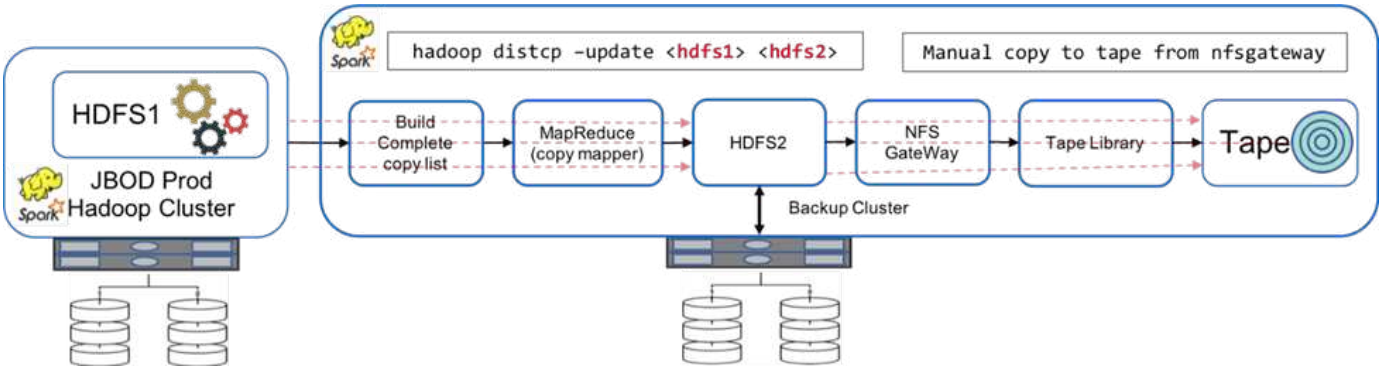
要求和挑战

此用例的主要要求和挑战包括：

- 软件向后兼容性：
  - 建议的备用备份解决方案应与生产 Hadoop 集群中当前正在运行的软件版本兼容。
- 为了满足承诺的 SLA ，建议的替代解决方案应实现极低的 RPO 和 RTO 。
- NetApp 备份解决方案创建的备份可以在本地构建于数据中心的 Hadoop 集群中使用，也可以在远程站点的灾难恢复位置运行的 Hadoop 集群中使用。
- 建议的解决方案必须经济高效。
- 建议的解决方案必须减少备份期间对当前正在运行的生产分析作业的性能影响。

客户的现有备份解决方案

下图显示了原始 Hadoop 原生备份解决方案。



生产数据通过中间备份集群保护到磁带：

- 通过运行 `hadoop distcp -update <hdfs1> <hdfs2>` 命令，可以将 HDFS1 数据复制到 HDFS2 。
- 备份集群充当 NFS 网关，数据通过 Linux `CP` 命令通过磁带库手动复制到磁带。

原始 Hadoop 原生备份解决方案的优势包括：

- 解决方案基于 Hadoop 原生命令，用户无需学习新过程。
- 解决方案采用行业标准架构和硬件。

原始 Hadoop 原生备份解决方案的缺点包括：

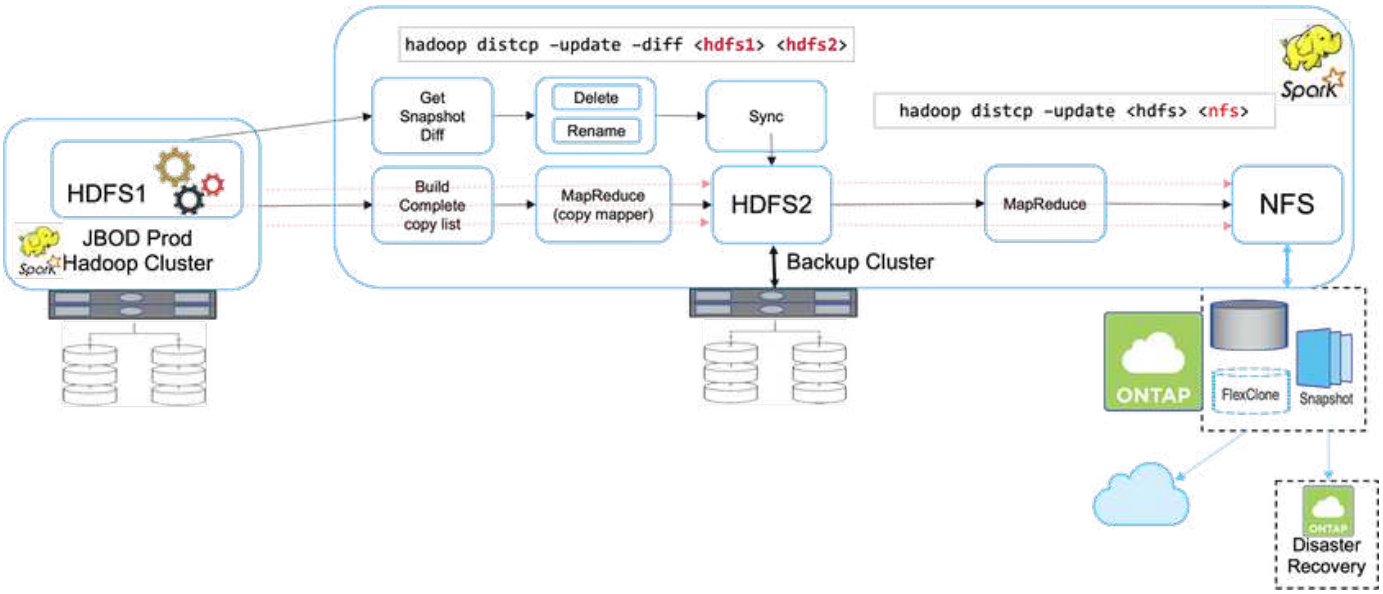
- 备份窗口时间过长超过 24 小时，因此生产数据容易受到影响。
- 备份期间集群性能显著下降。
- 复制到磁带是一个手动过程。
- 备份解决方案在所需的硬件和手动过程所需的工时方面成本高昂。

备份解决方案

根据这些挑战和要求，并考虑到现有备份系统，建议了三种可能的备份解决方案。以下各小节分别介绍了这三种不同的备份解决方案，分别标记为解决方案 A 到解决方案 C

解决方案 **A**

在解决方案A中、备份Hadoop集群会将二级备份发送到NetApp NFS存储系统、从而无需使用磁带、如下图所示。



解决方案 A 的详细任务包括：

- 生产 Hadoop 集群的 HDFS 中包含需要保护的客户分析数据。
- 具有 HDFS 的备份 Hadoop 集群充当数据的中间位置。在生产和备份 Hadoop 集群中，只需一组磁盘（ JBOD ）即可为 HDFS 提供存储。
- 运行 `hadoop distcp - update - diff <hdfs1> <hdfs2>` 命令，保护 Hadoop 生产数据，使其从

生产集群 HDFS 传输到备份集群 HDFS 。

> Hadoop 快照用于保护数据从生产环境传输到备份 Hadoop 集群。

- NetApp ONTAP 存储控制器可提供一个 NFS 导出卷，该卷会配置到备份 Hadoop 集群。
- 运行 `Hadoop distcp` 命令利用Mapreduce"和多个映射程序、可以保护分析数据从Hadoop集群备份到NFS。

  将数据存储在 NetApp 存储系统的 NFS 中后，将使用 NetApp Snapshot ， SnapRestore 和 FlexClone 技术根据需要备份，还原和复制 Hadoop 数据。

> 使用 SnapMirror 技术，可以将 Hadoop 数据保护到云端以及灾难恢复位置。

解决方案 A 的优势包括：

- Hadoop 生产数据受备份集群保护。
- HDFS 数据通过 NFS 进行保护，从而可以保护云和灾难恢复位置。
- 通过将备份操作卸载到备份集群来提高性能。
- 无需手动执行磁带操作
- 支持通过 NetApp 工具实现企业管理功能。
- 只需对现有环境进行极少的更改。
- 是一种经济高效的解决方案。

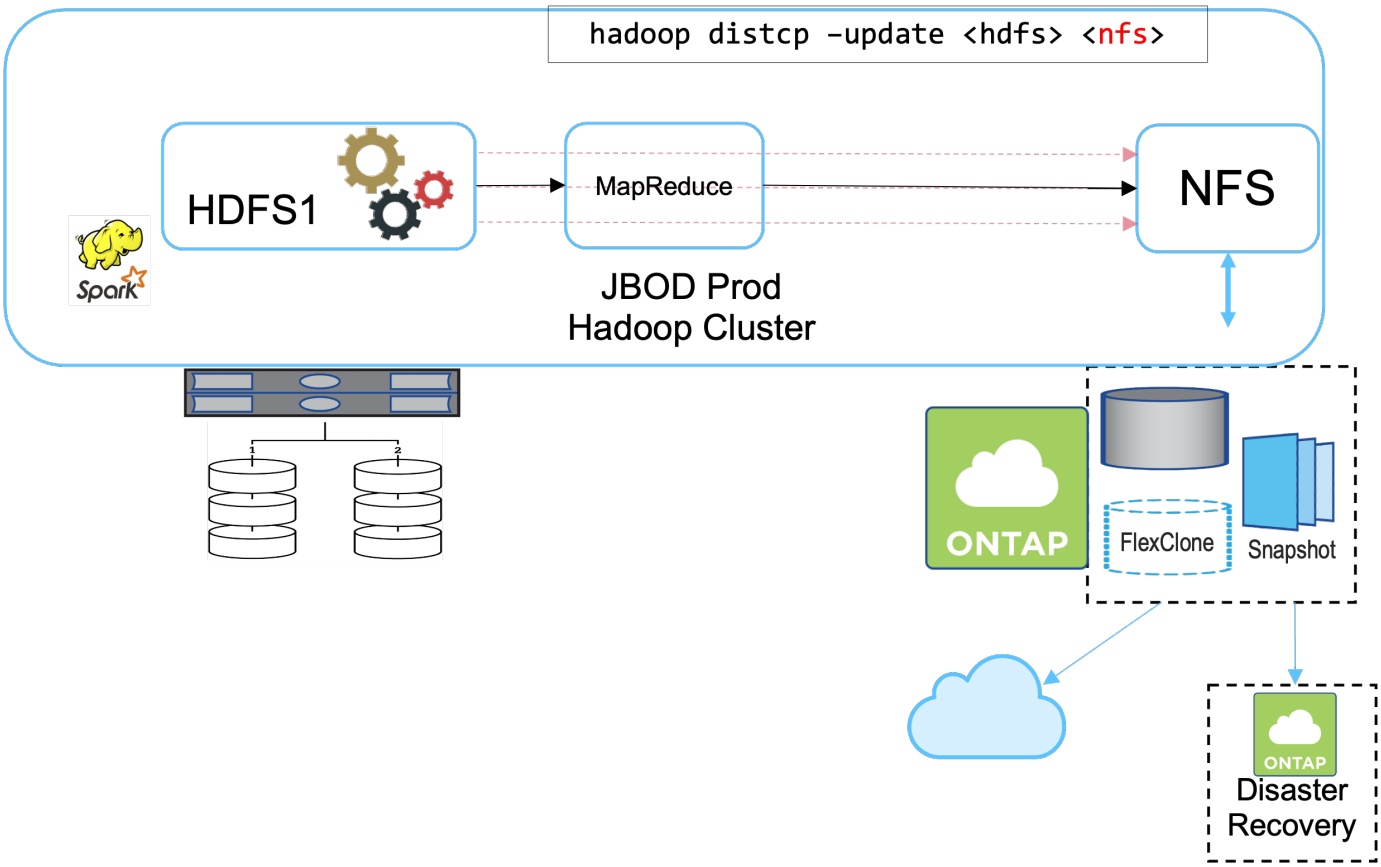
此解决方案的缺点是，它需要一个备份集群和额外的映射器来提高性能。

由于解决方案 A 的简单性，成本和整体性能，客户最近部署了它。

在此解决方案中，可以使用 ONTAP 中的 SAN 磁盘，而不是 JBOD 。此选项会将备份集群存储负载卸载到 ONTAP ；但是，缺点是需要 SAN 网络结构交换机。

解决方案 **B**

解决方案B会将NFS卷添加到生产Hadoop集群中、从而无需备份Hadoop集群、如下图所示。



解决方案 B 的详细任务包括：

- NetApp ONTAP 存储控制器可为生产 Hadoop 集群配置 NFS 导出。

  Hadoop本机 `hadoop distcp` 命令可保护Hadoop数据从生产集群HDFS到NFS。

- 将数据存储在 NetApp 存储系统的 NFS 中后，将使用 Snapshot ， SnapRestore 和 FlexClone 技术根据需要备份，还原和复制 Hadoop 数据。

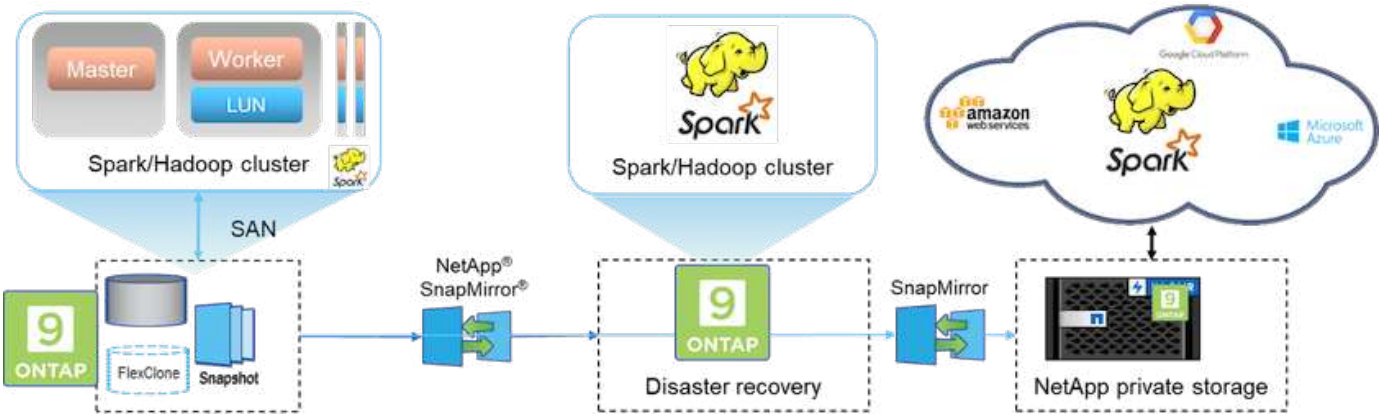
解决方案 B 的优势包括：

- 生产集群针对备份解决方案进行了少许修改，从而简化了实施并降低了额外的基础架构成本。
- 备份操作不需要备份集群。
- HDFS 生产数据在转换为 NFS 数据时会受到保护。
- 解决方案支持通过 NetApp 工具执行企业管理功能。

此解决方案的缺点是它在生产集群中实施，这可能会在生产集群中添加其他管理员任务。

解决方案 **C**

在解决方案 C 中， NetApp SAN 卷会直接配置到 Hadoop 生产集群中以用于 HDFS 存储，如下图所示。



解决方案 C 的详细步骤包括：

- NetApp ONTAP SAN 存储在生产 Hadoop 集群上配置为用于 HDFS 数据存储。
- NetApp Snapshot 和 SnapMirror 技术用于备份生产 Hadoop 集群中的 HDFS 数据。
- 在 Snapshot 副本备份过程中， Hadoop/Spark 集群的生产不会对性能产生影响，因为备份位于存储层。

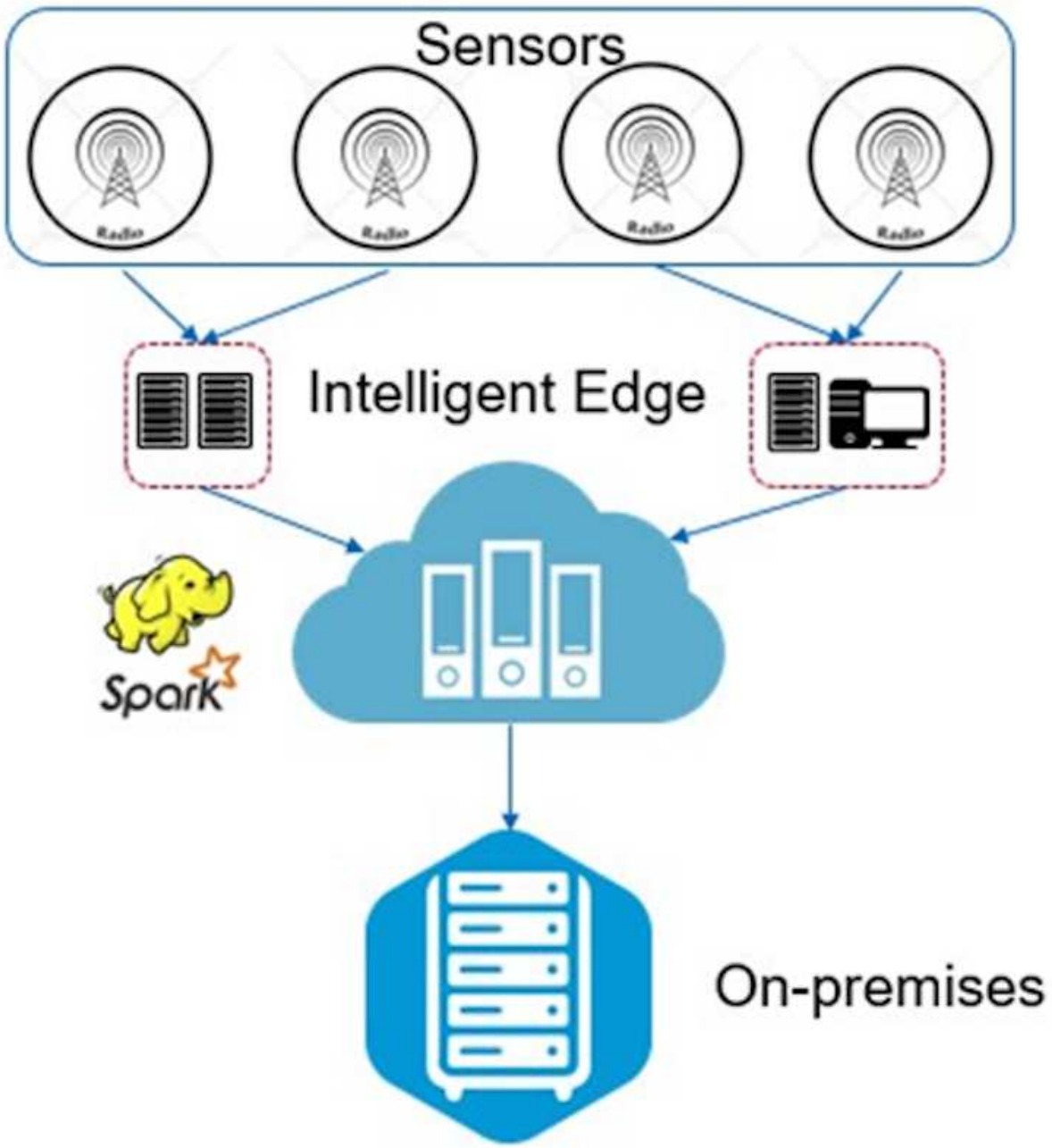
> 无论数据大小如何， Snapshot 技术均可在数秒内完成备份。

解决方案 C 的优势包括：

- 可以使用 Snapshot 技术创建节省空间的备份。
- 支持通过 NetApp 工具实现企业管理功能。

用例 2： 从云到内部环境的备份和灾难恢复

此使用情形基于广播客户，该客户需要将基于云的分析数据备份到其内部数据中心，如下图所示。



场景

在这种情况下， IoT 传感器数据会载入云中，并使用 AWS 中的开源 Apache Spark 集群进行分析。需要将处理后的数据从云备份到内部环境。

要求和挑战

此用例的主要要求和挑战包括：

- 启用数据保护不会发生原因对云中的生产 Spark 或 Hadoop 集群产生任何性能影响。
- 云传感器数据需要以高效安全的方式移动和保护到内部。
- 可以在不同条件下灵活地将数据从云传输到内部环境，例如按需，瞬时以及在集群负载较低的情况下。

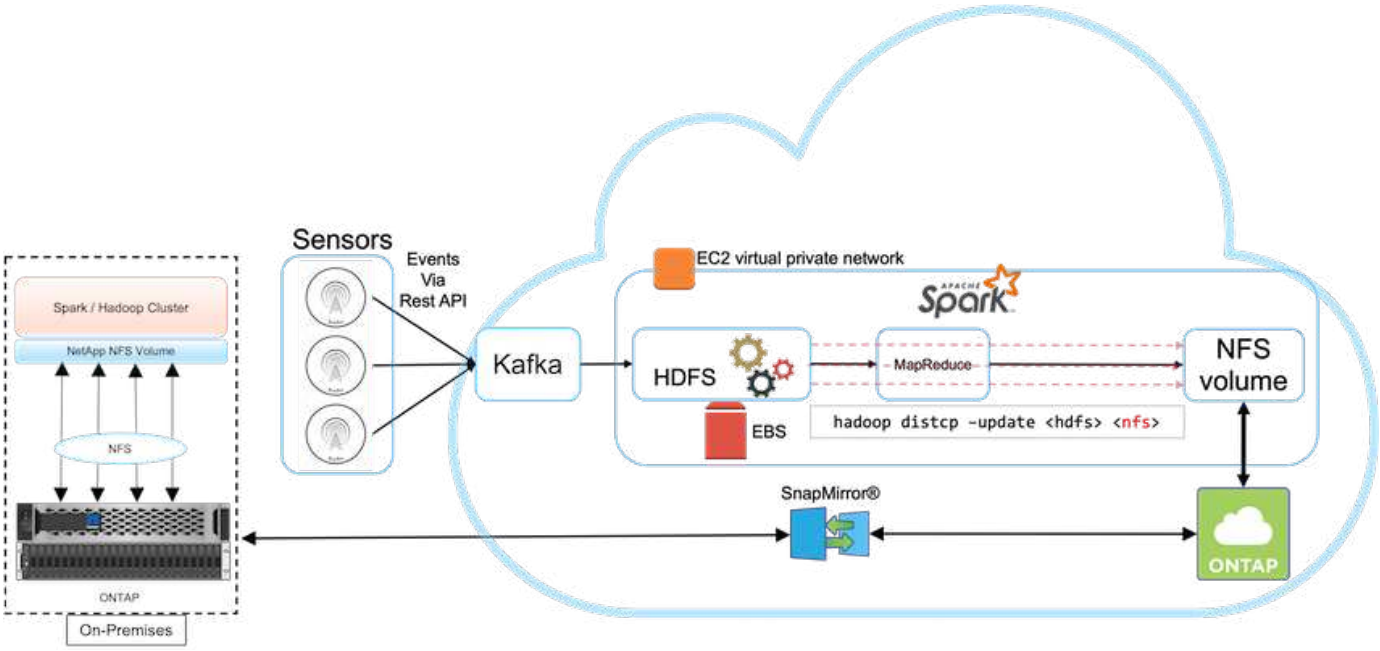
解决方案

客户使用 AWS Elastic Block Store （ EBS ）作为其 Spark 集群 HDFS 存储，通过 Kafka 从远程传感器接收和载入数据。因此， HDFS 存储充当备份数据的源。

为了满足这些要求， NetApp ONTAP 云部署在 AWS 中，并创建一个 NFS 共享作为 Spark 或 Hadoop 集群的备份目标。

创建NFS共享后、将数据从HDFS EBS存储复制到ONTAP NFS共享中。数据驻留在 ONTAP 云的 NFS 中后，可以根据需要使用 SnapMirror 技术将数据从云镜像到内部存储，从而实现安全高效的方式。

此图显示了从云到内部解决方案的备份和灾难恢复。



用例 3： 对现有 Hadoop 数据启用 DevTest

在此使用情形中，客户要求在同一数据中心和远程位置快速高效地基于包含大量分析数据以用于 DevTest 和报告目的的现有 Hadoop 集群构建新的 Hadoop/Spark 集群。

场景

在这种情况下，多个 Spark 或 Hadoop 集群是通过在内部以及灾难恢复位置实施大型 Hadoop 数据湖而构建的。

要求和挑战

此用例的主要要求和挑战包括：

- 创建多个 Hadoop 集群以实现 DevTest ， QA 或任何其他需要访问相同生产数据的目的。此处的挑战是，以节省空间的方式瞬时克隆多个非常大的 Hadoop 集群。
- 将 Hadoop 数据同步到 DevTest 和报告团队，以提高运营效率。
- 使用相同的凭据在生产集群和新集群之间分发 Hadoop 数据。

- 使用计划策略高效创建 QA 集群，而不会影响生产集群。

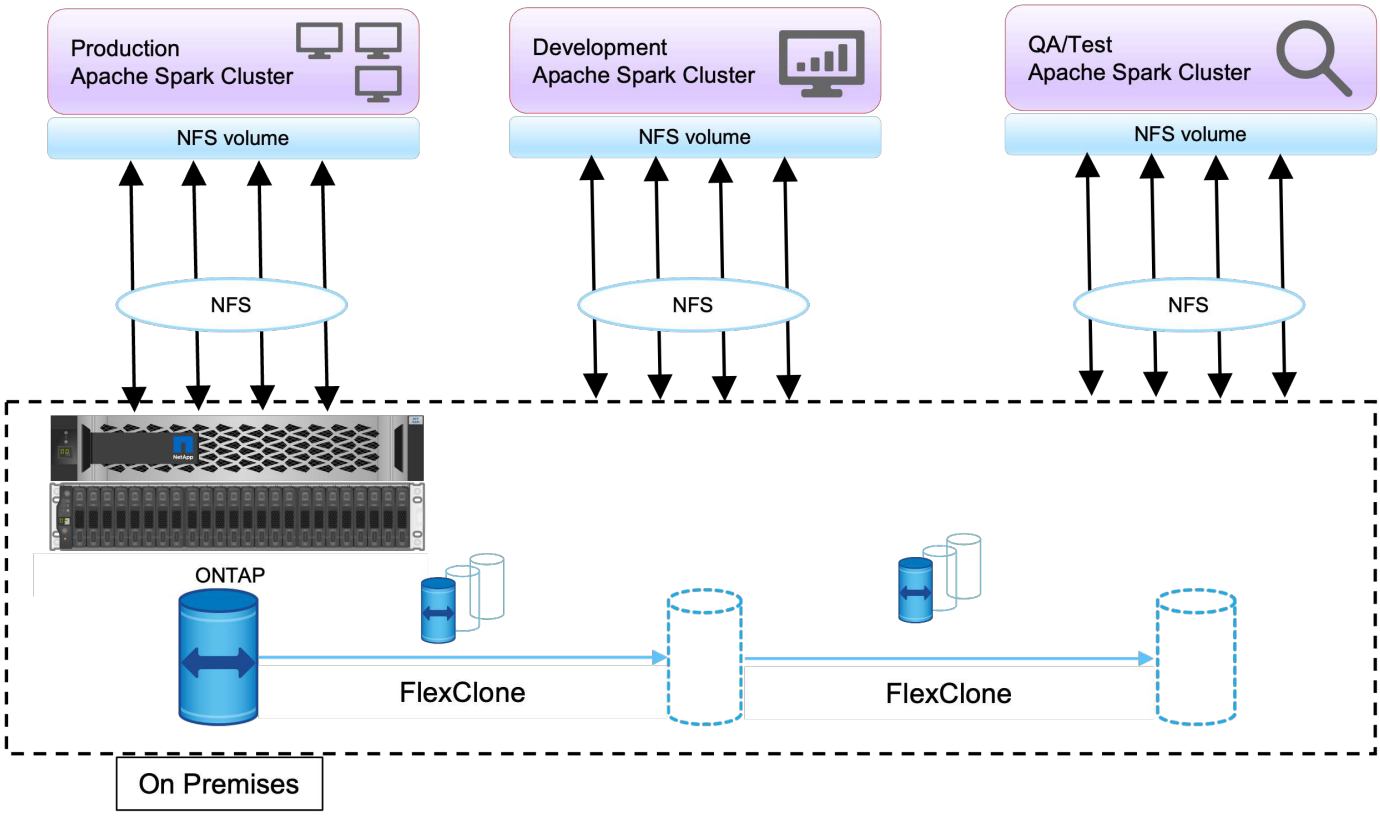
解决方案

FlexClone 技术用于问题解答上述要求。FlexClone 技术是 Snapshot 副本的读 / 写副本。它会从父 Snapshot 副本数据读取数据，并且只会为新的 / 修改的块占用额外空间。速度快，节省空间。

首先，使用 NetApp 一致性组创建现有集群的 Snapshot 副本。

NetApp System Manager 或存储管理员提示符中的 Snapshot 副本。一致性组 Snapshot 副本是应用程序一致的组 Snapshot 副本， FlexClone 卷是根据一致性组 Snapshot 副本创建的。值得一提的是， FlexClone 卷会继承父卷的 NFS 导出策略。创建 Snapshot 副本后，必须安装一个新的 Hadoop 集群以用于 DevTest 和报告目的，如下图所示。从新Hadoop集群克隆的NFS卷可访问NFS数据。

此图显示了用于 DevTest 的 Hadoop 集群。



用例 4： 数据保护和多云连接

此使用情形适用于负责为客户的大数据分析数据提供多云连接的云服务合作伙伴。

场景

在这种情况下，在 AWS 中从不同来源接收的物联网数据存储在 NPS 的中央位置。NPS 存储连接到 AWS 和 Azure 中的 Spark 或 Hadoop 集群，从而支持在多个云中运行的大数据分析应用程序访问相同的数据。

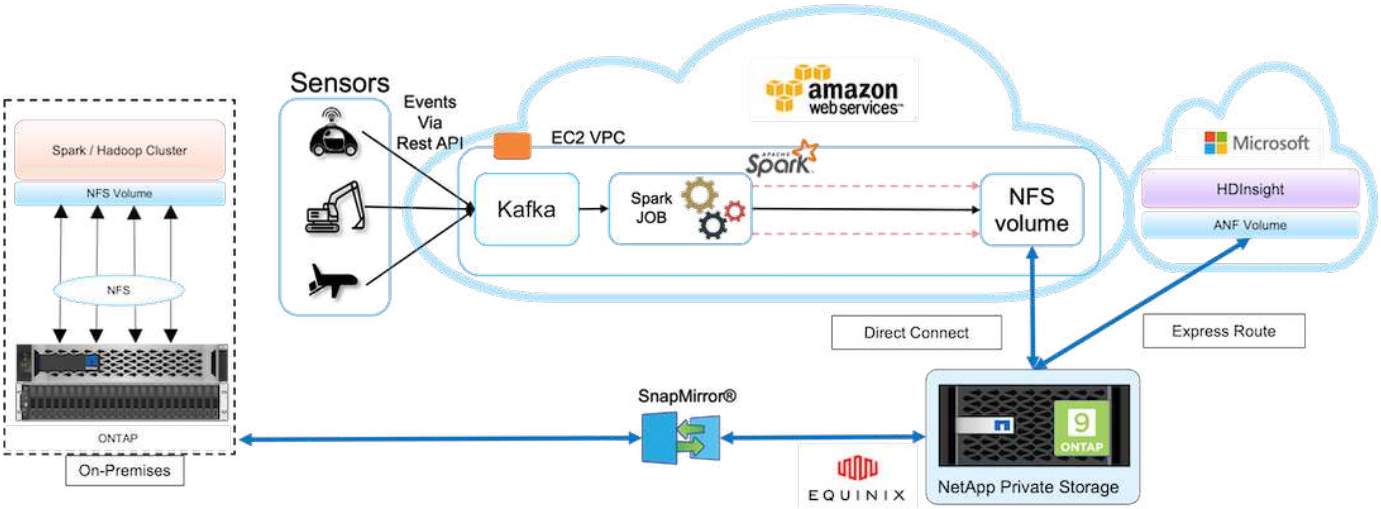
要求和挑战

此用例的主要要求和挑战包括：

- 客户希望使用多个云对相同数据运行分析作业。
- 必须通过不同的传感器和中心从内部和云等不同来源接收数据。
- 解决方案必须高效且经济高效。
- 主要挑战是构建一个经济高效的解决方案，以便在内部和不同云之间提供混合分析服务。

解决方案

此图显示了数据保护和多云连接解决方案。



如上图所示，来自传感器的数据会通过 Kafka 流式传输并输入到 AWS Spark 集群中。数据存储在 NPS 中的 NFS 共享中， NPS 位于 Equinix 数据中心内的云提供商之外。由于NetApp NPS分别通过Direct Connect和Express Route连接到Amazon AWS和Microsoft Azure、因此客户可以从Amazon和AWS分析集群访问NFS数据。这种方法解决了跨多个超大规模云提供商进行云分析的问题。

因此，由于内部和 NPS 存储均运行 ONTAP 软件，因此 SnapMirror 可以将 NPS 数据镜像到内部集群，从而在内部和多个云之间提供混合云分析。

为了获得最佳性能， NetApp 通常建议使用多个网络接口和直接连接 / 快速路由从云实例访问数据。

用例 5： 加快分析工作负载的速度

在这种情况下，一家大型金融服务和投资银行的分析平台使用 NetApp NFS 存储解决方案进行了现代化改造，从而显著改进了资产管理和定量业务部门的投资风险和衍生产品分析。

场景

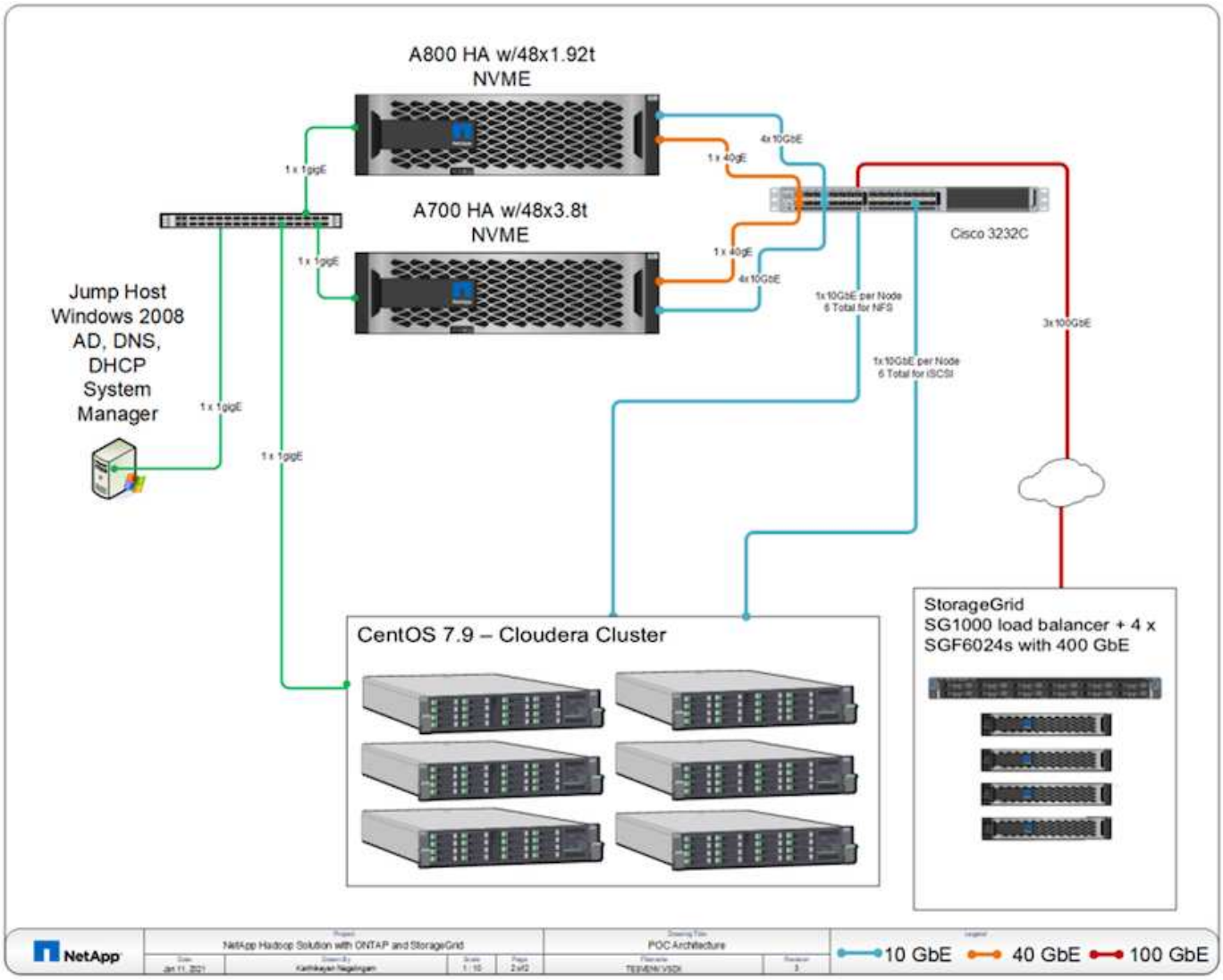
在客户的现有环境中，用于分析平台的 Hadoop 基础架构利用了 Hadoop 服务器的内部存储。由于 JBOD 环境的专有性，组织内的许多内部客户无法利用其蒙特卡罗量化模型，这种模拟依赖于重复的实时数据样本。无法以最佳方式了解市场波动的不确定性影响，这对量化资产管理业务单位不利。

要求和挑战

该银行的量化业务部门需要一种高效的预测方法来实现准确，及时的预测。为此，该团队认识到需要对基础架构进行现代化改造，减少现有 I/O 等待时间并提高 Hadoop 和 Spark 等分析应用程序的性能，以便高效模拟投资模型，衡量潜在收益并分析风险。

解决方案

客户的现有 Spark 解决方案具有 JBOD 。然后，利用 NetApp ONTAP ， NetApp StorageGRID 和从 MinIO 网关到 NFS ，减少了银行的量化财务团队对评估潜在收益和风险的投资模型进行模拟和分析的 I/O 等待时间。此图显示了采用 NetApp 存储的 Spark 解决方案。



如上图所示，我们部署了 AFF A800 ， A700 系统和 StorageGRID ，以便在具有 Spark 的六节点 Hadoop 集群中通过 NFS 和 S3 协议访问 parquet 文件，并为数据分析操作提供了 YARN 和 Hive 元数据服务。

客户旧环境中的直连存储（ DAS ）解决方案在独立扩展计算和存储方面存在缺点。借助 NetApp ONTAP 解决方案 for Spark ，该银行的财务分析业务部门能够将存储与计算分离，并根据需要无缝地更有效地提供基础架构资源。

通过将 ONTAP 与 NFS 结合使用，计算服务器 CPU 几乎可以完全用于 Spark SQL 作业， I/O 等待时间缩短了近 70% ，从而为 Spark 工作负载提供了更好的计算能力和性能提升。随后，随着 CPU 利用率的提高，客户还

可以利用 GPUDirect 等 GPU 进一步实现平台现代化。此外， StorageGRID 还为 Spark 工作负载提供了低成本存储选项，而 MinIO 网关则可通过 S3 协议安全访问 NFS 数据。对于云中的数据， NetApp 建议使用 Cloud Volumes ONTAP ， Azure NetApp Files 和 NetApp Cloud Volumes Service 。

结论

本节概述了 NetApp 为满足各种 Hadoop 数据保护要求而提供的使用情形和解决方案。通过使用 NetApp 提供支持的 Data Fabric ，客户可以：

- 利用 NetApp 丰富的数据管理功能并与 Hadoop 原生工作流集成，灵活选择合适的数据保护解决方案。
- 将 Hadoop 集群备份窗口时间缩短近 70% 。
- 消除 Hadoop 集群备份所带来的任何性能影响。
- 同时提供从不同云提供商到单个分析数据源的多云数据保护和数据访问。
- 使用 FlexClone 技术快速创建节省空间的 Hadoop 集群副本。

从何处查找追加信息

要了解有关本文档中所述信息的更多信息，请参见以下文档和 / 或网站：

- NetApp 大数据分析解决方案

  "https://www.netapp.com/us/solutions/applications/big-data-analytics/index.aspx"

- 采用 NetApp 存储的 Apache Spark 工作负载

  https://www.netapp.com/pdf.html?item=/media/26877-nva-1157-deploy.pdf

- 适用于 Apache Spark 的 NetApp 存储解决方案

  "https://www.netapp.com/media/16864-tr-4570.pdf"

- NetApp 支持的 Data Fabric 上的 Apache Hadoop

  "https://www.netapp.com/media/16877-tr-4529.pdf"

致谢

- NetApp ANZ 维多利亚地区销售代表 Paul Burland
- NetApp 业务开发经理 Hoseb Dermanilian
- NetApp MPSG 总监 Lee Dorrier
- NetApp ANZ 维多利亚区 SE 系统工程师 David Thiessen

版本历史记录

version	**Date**	文档版本历史记录
版本 1.0	2018 年 1 月	初始版本。

version	Date	文档版本历史记录
版本 2.0	2021年10月	更新了用例 5： 加快分析工作负载的速度
版本 3.0	2023年11月	已删除NIPAM详细信息

版权信息

版权所有 © 2024 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本文档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。 美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 http://www.netapp.com/TM 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。