Kafka与NetApp ONTAP 存储控制器相结合

NetApp Solutions

NetApp
April 12, 2024

This PDF was generated from https://docs.netapp.com/zh-cn/netapp-solutions/data-analytics/kafka-ontap-overview.html on April 12, 2024. Always check docs.netapp.com for the latest.

目录

Kafka与NetApp ONTAP 存储控制器相结合 . . . 1
TR-4941：与NetApp ONTAP 存储控制器相结合 . . . 1
解决方案 . . . 1
技术概述 . . . 2
融合性能验证 . . . 7
使用生产用工作负载生成器进行性能测试 . . . 10
性能最佳实践准则 . . . 11
结论 . . . 12

Kafka与NetApp ONTAP 存储控制器相结合

TR-4941：与NetApp ONTAP 存储控制器相结合

Karthikeyan Nagalingam、Joe Scott、NetApp Rankesh Kumar、Confluent

要使Confluent平台更具可扩展性和弹性、IT必须能够快速扩展和平衡工作负载。分层存储可减轻这种操作负担、从而使在Confluent中存储海量数据变得易于管理。

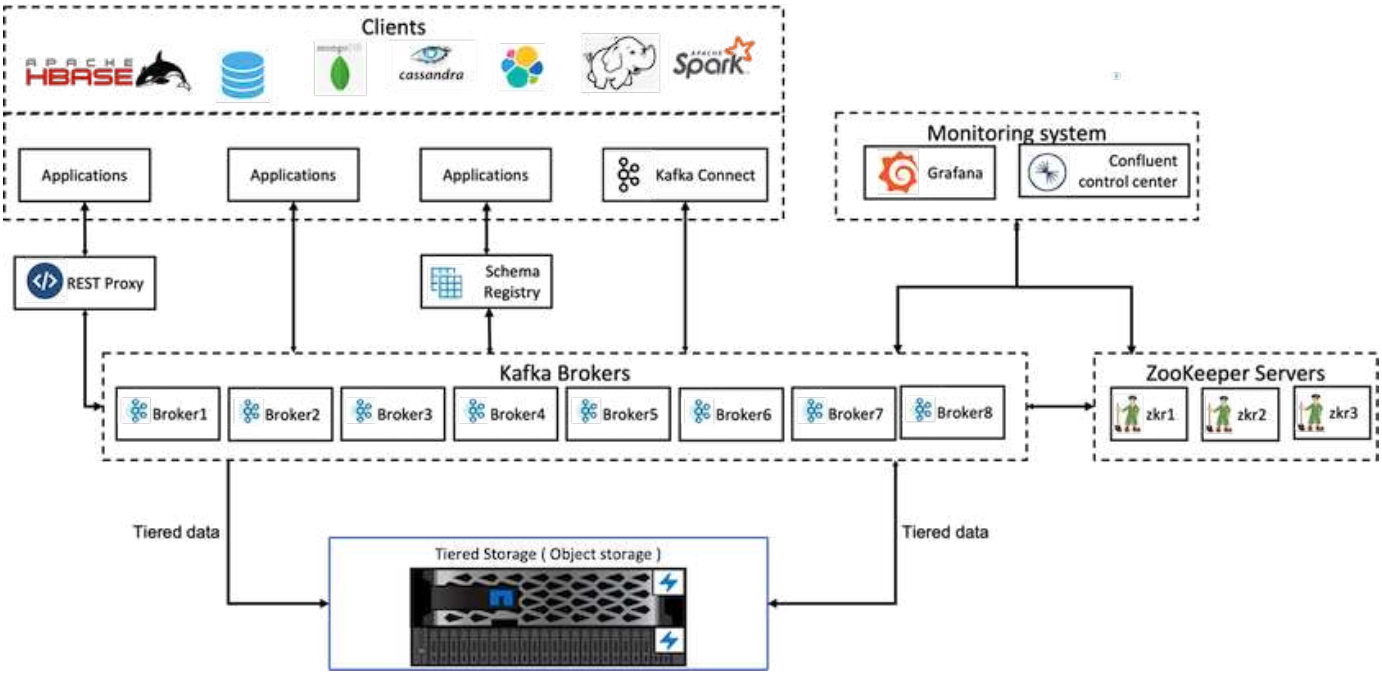
基本理念是将数据存储与数据处理分离、这样可以更轻松地独立扩展每个存储。

NetApp ONTAP 数据管理软件集行业领先的创新技术于一身、无论数据位于何处、均可为Confluent提供诸多优势。

本文档概述了使用分层存储基准测试套件的NetApp ONTAP 上的Confluent平台的性能基准测试。

解决方案

由ONTAP 提供支持的Confuent和NetApp AFF A900存储控制器是专为数据流设计的分布式系统。两者均可水平扩展、容错、并在负载下提供出色的性能。它们通过数据精简技术在分布式数据流和流处理方面相辅相成、并可降低存储成本、从而最大限度地减少数据占用空间。AFF A900存储控制器可提供出色的性能、同时允许分离计算和数据存储资源。这样可以简化系统管理并独立扩展资源。



解决方案架构详细信息

本节介绍在使用NetApp ONTAP 进行分层存储的Confluent Platform部署中用于性能验证的硬件和软件。下表介绍了解决方案 架构和基本组件。

平台组件	环境配置
Confuent Platform 6.2版	• 3个Zookepers • 8个代理服务器 • 5个工具服务器 • 1个Grafana • 1个控制中心
所有节点上的操作系统	Linux （ Ubuntu 18.04 ）
适用于温分段的NetApp ONTAP	• 1个AFF A900高可用性(HA)对 • 4 个 24 x 800 SSD • S3 协议 • 100GbE
15 台 Fujitsu PRIMERGY RX2540 服务器	• 2个CPU；总共16个物理核心 • Intel Xeon • 256 GB物理内存 • 100GbE双端口

技术概述

本节介绍此解决方案中使用的技术。

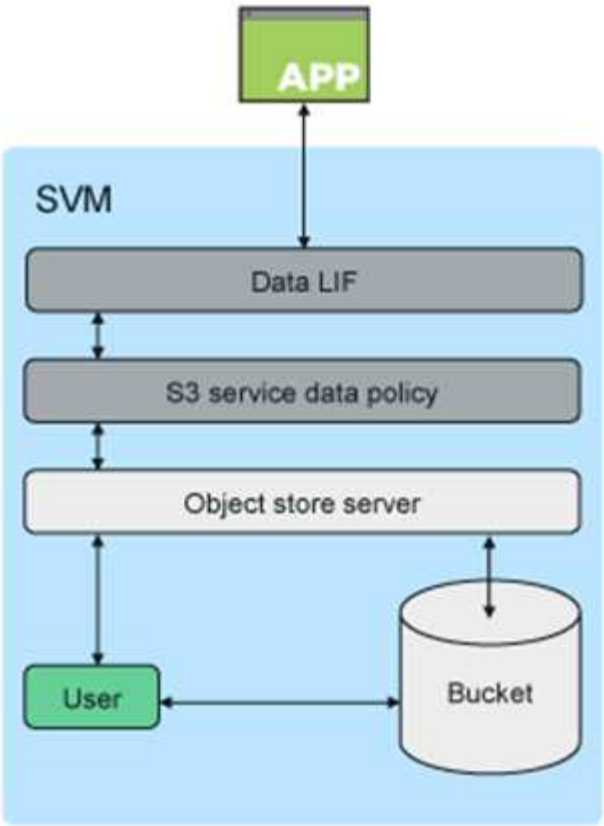
NetApp ONTAP 存储控制器

NetApp ONTAP 是一款高性能企业级存储操作系统。

NetApp ONTAP 9.8引入了对Amazon Simple Storage Service (S3) API的支持。ONTAP 支持部分Amazon Web Services (AWS) S3 API操作、并允许在云提供商(AWS、Azure和GCP)和内部环境中将数据表示为基于ONTAP的系统中的对象。

NetApp StorageGRID 软件是用于对象存储的旗舰级NetApp解决方案。ONTAP 通过在边缘提供载入和预处理点、扩展由NetApp提供支持的对象数据数据数据网络结构以及提高NetApp产品组合的价值、对StorageGRID 进行了补充。

可以通过授权的用户和客户端应用程序访问S3存储分段。下图显示了访问S3存储分段的应用程序。



主要用例

支持S3 API的主要目的是在ONTAP 上提供对象访问。ONTAP 统一存储架构现在支持文件(NFS和SMB)、块(FC和iSCSI)和对象(S3)。

原生 S3应用程序

越来越多的应用程序能够使用S3利用ONTAP 支持对象访问。虽然非常适合高容量归档工作负载、但原生 S3应用程序对高性能的需求正在快速增长、其中包括：

- 分析
- 人工智能
- 边缘到核心载入
- 机器学习

现在、客户可以使用熟悉的易管理性工具(如ONTAP System Manager)快速配置高性能对象存储、以便在ONTAP 中进行开发和操作、同时充分利用ONTAP 的存储效率和安全性。

FabricPool 端点

从ONTAP 9.8开始、FabricPool 支持在ONTAP 中分层到分段、从而可以进行ONTAP到ONTAP分层。对于希望将现有FAS 基础架构重新用作对象存储端点的客户来说、这是一个绝佳的选择。

FabricPool 支持通过两种方式分层到ONTAP：

- *本地集群分层。*使用集群LIF将非活动数据分层到位于本地集群上的存储分段。
- *远程集群分层。*非活动数据将采用与传统FabricPool 云层类似的方式分层到远程集群上的存储分层、方法是在FabricPool 客户端上使用IC LIF和在ONTAP 对象存储上使用数据LIF。

如果您需要在现有集群上使用 S3 功能，而无需额外的硬件和管理，则 ONTAP S3 是合适的。对于300 TB以上的部署、NetApp StorageGRID 软件仍然是NetApp对象存储的旗舰级解决方案。使用ONTAP 或StorageGRID作为云层时、不需要FabricPool 许可证。

适用于**Confluent**分层存储的**NetApp ONTAP**

每个数据中心都需要保持业务关键型应用程序持续运行、并确保重要数据的可用性和安全性。全新的NetApp AFF A900系统采用ONTAP 企业版软件和高弹性设计。我们全新的快如闪电的NVMe存储系统可消除任务关键型运营中断、最大限度地降低性能调整、并保护您的数据免受勒索软件攻击。

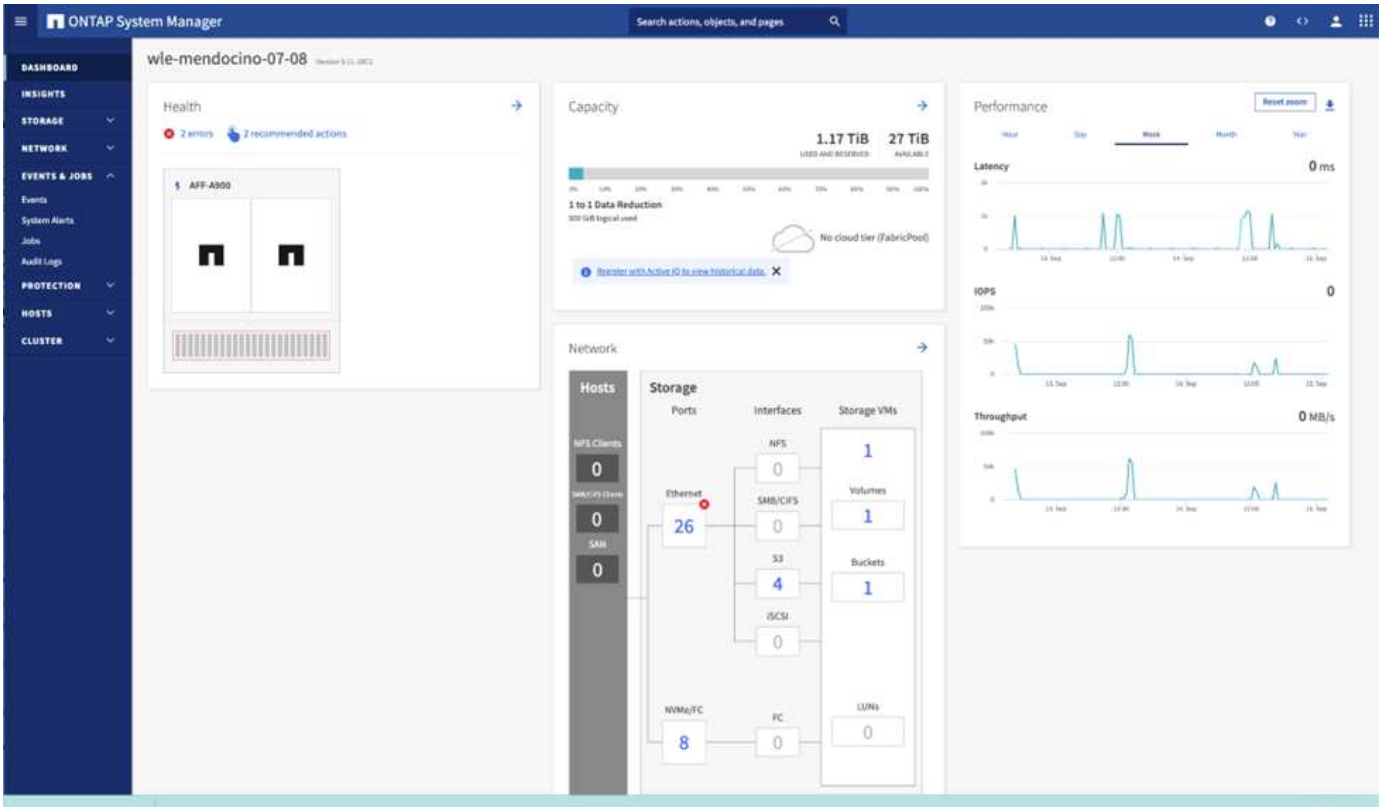
从初始部署到扩展Confluent集群、您的环境需要快速适应不会对业务关键型应用程序造成中断的变化。ONTAP企业级数据管理、服务质量(Quality of Service、QoS)和性能支持您规划和适应您的环境。

将NetApp ONTAP 和Confluent分层存储结合使用、可将ONTAP 用作横向扩展存储目标、从而简化Apache Kafka集群的管理、并可为Confluent独立扩展计算和存储资源。

ONTAP S3服务器基于ONTAP 成熟的横向扩展存储功能构建。通过扩展S3存储分段、将新添加的节点添加到ONTAP 集群、可以无缝扩展ONTAP 集群。

使用**ONTAP System Manager**进行简单管理

ONTAP System Manager是一个基于浏览器的图形界面、可用于在一个管理平台中跨全球分布位置配置、管理和监控ONTAP 存储控制器。



您可以使用System Manager和ONTAP 命令行界面配置和管理ONTAP S3。当您启用S3并使用System Manager创建存储分段时、ONTAP 会为简化的配置提供最佳实践默认值。如果您从CLI配置S3服务器和存储分段、则仍可根据需要使用System Manager对其进行管理、反之亦然。

使用 System Manager 创建 S3 存储分段时， ONTAP 会配置系统上可用性最高的默认性能服务级别。例如、在AFF 系统上、默认设置为"Extreme"。性能服务级别是预定义的自适应QoS策略组。您可以指定自定义 QoS 策略组，也可以不指定策略组，而不指定默认服务级别之一。

预定义的自适应QoS策略组包括以下内容：

- 至尊。*用于需要最低延迟和最高性能的应用程序。
- *性能。*用于性能需求和延迟适中的应用程序。
- 值。用于吞吐量和容量比延迟更重要的应用程序。
- *自定义。*指定自定义QoS策略或不指定QoS策略。

如果选择 * 用于分层 * ，则不会选择任何性能服务级别，系统会尝试为分层数据选择具有最佳性能的低成本介质。

ONTAP 会尝试在磁盘最合适的本地层上配置此存储分段，以满足所选的服务级别。但是，如果需要指定要包含在存储分段中的磁盘，请考虑通过指定本地层（聚合）从 CLI 配置 S3 对象存储。如果您通过 CLI 配置 S3 服务器，则仍可根据需要使用 System Manager 对其进行管理。

如果您希望能够指定用于存储分段的聚合，则只能使用命令行界面来执行此操作。

两者结合

Confuent Platform 是一个全面的数据流平台，可让您轻松地以持续的实时流的形式访问，存储和管理数据。Confluent 由 Apache Kafka 的原始创建者构建，通过企业级功能扩展了 Kafka 的优势，同时消除了 Kafka 的管理或监控负担。如今、《财富》100强企业中有80%以上的企业采用数据流技术、大多数企业都使用Confluent。

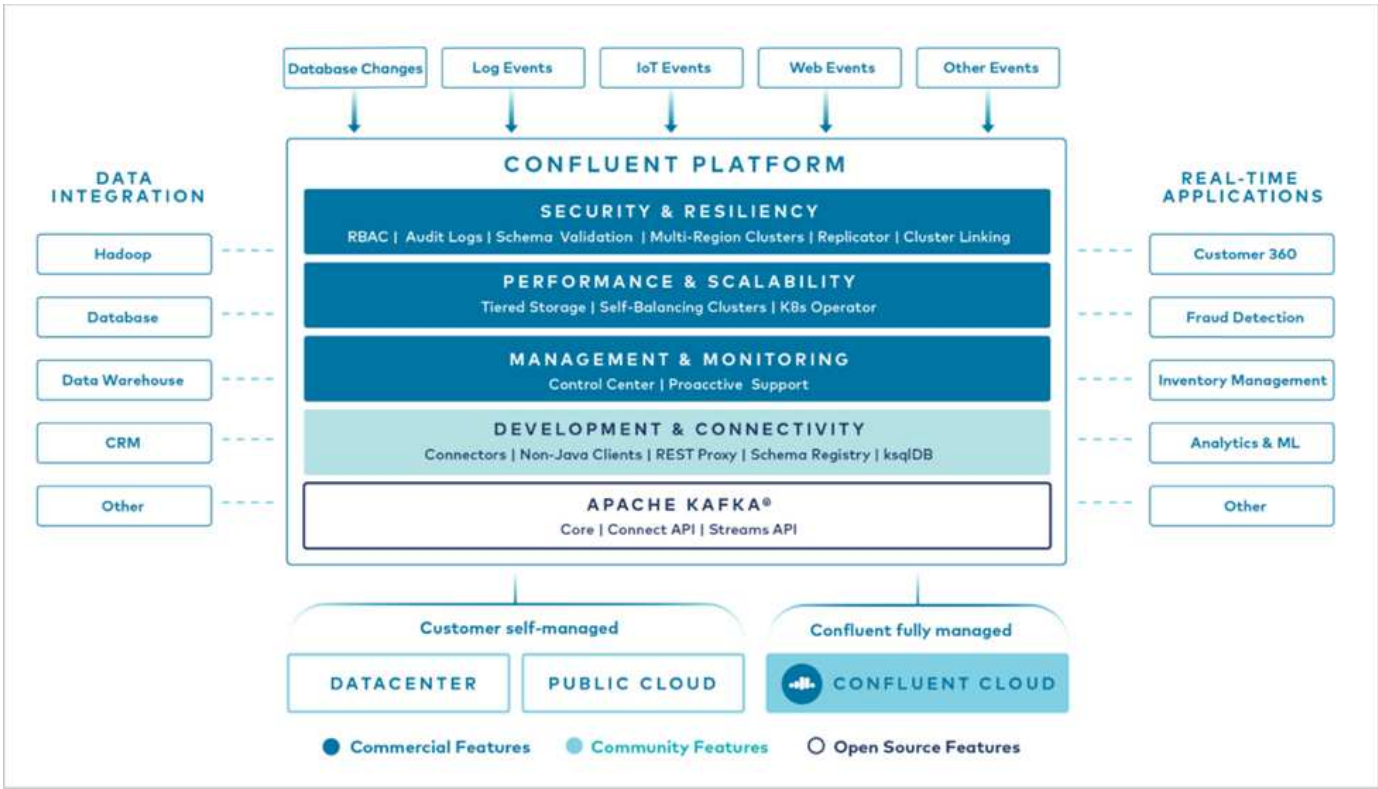
为什么选择 **Confluent** ？

通过将历史数据和实时数据集成到一个统一的中央真相来源中， Confluent 可以轻松构建一个全新的现代化事件驱动型应用程序类别，获得通用数据管道，并充分扩展性，性能和可靠性，释放出强大的新用例。

Confluent 的用途是什么？

借助整合平台，您可以专注于如何从数据中获得业务价值，而不是担心底层机制，例如如何在不同系统之间传输或集成数据。具体而言， Confluent Platform 可简化将数据源连接到 Kafka 的过程，构建流式应用程序，以及保护，监控和管理 Kafka 基础架构。如今、Consfluent Platform已广泛用于各行各业的各种用例、从金融服务、全渠道零售和自动驾驶汽车到欺诈检测、微服务和物联网。

下图显示了Confluent Platform的组件。



Confluent事件流技术概述

Confluent Platform 的核心是 "Kafka"最受欢迎的开源分布式流式平台。Kafka的主要功能包括：

- 发布并订阅记录流。
- 以容错方式存储记录流。
- 处理记录流。

即装即用的 Confluent 平台还包括架构注册表， REST 代理，总共 100 多个预构建的 Kafka 连接器和 ksqlDB 。

Confluent平台企业功能概述

- 流畅控制中心。一种基于UI的系统、用于管理和监控Kafka。您可以通过它轻松管理 Kafka Connect ，以及创建，编辑和管理与其他系统的连接。
- * Kubernetes 的 Confluent 。 * Kubernetes 的 Confluent 是 Kubernetes 的操作员。Kubernetes 操作员通过为特定平台应用程序提供独特的功能和要求，扩展了 Kubernetes 的业务流程功能。对于 Confluent Platform ，这包括大幅简化 Kubernetes 上 Kafka 的部署流程，并自动执行典型的基础架构生命周期任务。
- *卡夫卡连接连接器。*连接器使用Kafka Connect API将Kafka连接到数据库、密钥值存储、搜索索引和文件系统等其他系统。Confluent Hub 提供可下载的连接器，用于最常用的数据源和数据池，包括这些连接器经过全面测试且受支持的版本以及 Confluent 平台。有关更多详细信息，请参见 "此处"。
- * 自平衡集群。 * 提供自动化负载平衡，故障检测和自我修复功能。它还支持根据需要添加或停用代理、而无需手动调整。
- * 流畅集群链接。 * 直接将集群连接在一起，并通过链路网桥将主题从一个集群镜像到另一个集群。集群链接可简化多数据中心，多集群和混合云部署的设置。
- *流畅自动数据平衡器。*监控集群中的代理数量、分区大小、分区数量和导数。它允许您在集群中移动数据

以创建均匀的工作负载，同时限制重新平衡流量，以便在重新平衡的同时最大限度地减少对生产工作负载的影响。

- * 流畅复制器。 * 使在多个数据中心维护多个 Kafka 集群变得比以往任何时候都更轻松。
- * 分层存储。 * 提供了使用您喜欢的云提供商存储大量 Kafka 数据的选项，从而减轻了运营负担并降低了成本。借助分层存储，您只能在需要更多计算资源时，才可以将数据保存在经济高效的对象存储和扩展代理上。
- * 流畅的 jms 客户端。 * 流畅平台包括适用于 Kafka 的与 jms 兼容的客户端。此 Kafka 客户端使用 Kafka 代理作为后端，实施了 Jms 1.1 标准 API 。如果旧版应用程序使用的是 jms ，并且您希望将现有的 jms 消息代理替换为 Kafka ，则此功能非常有用。
- * 流畅的 MQT 代理。 * 提供了一种从 MQT 设备和网关直接向 Kafka 发布数据的方法，而无需在中间使用 MQT 代理。
- * 流畅安全插件。 * 流畅安全插件用于为各种流畅平台工具和产品添加安全功能。目前，可以为 Confluent REST 代理提供一个插件，用于对传入请求进行身份验证，并将经过身份验证的主体传播到 Kafka 请求。这样， Confluent REST 代理客户端便可利用 Kafka 代理的多租户安全功能。

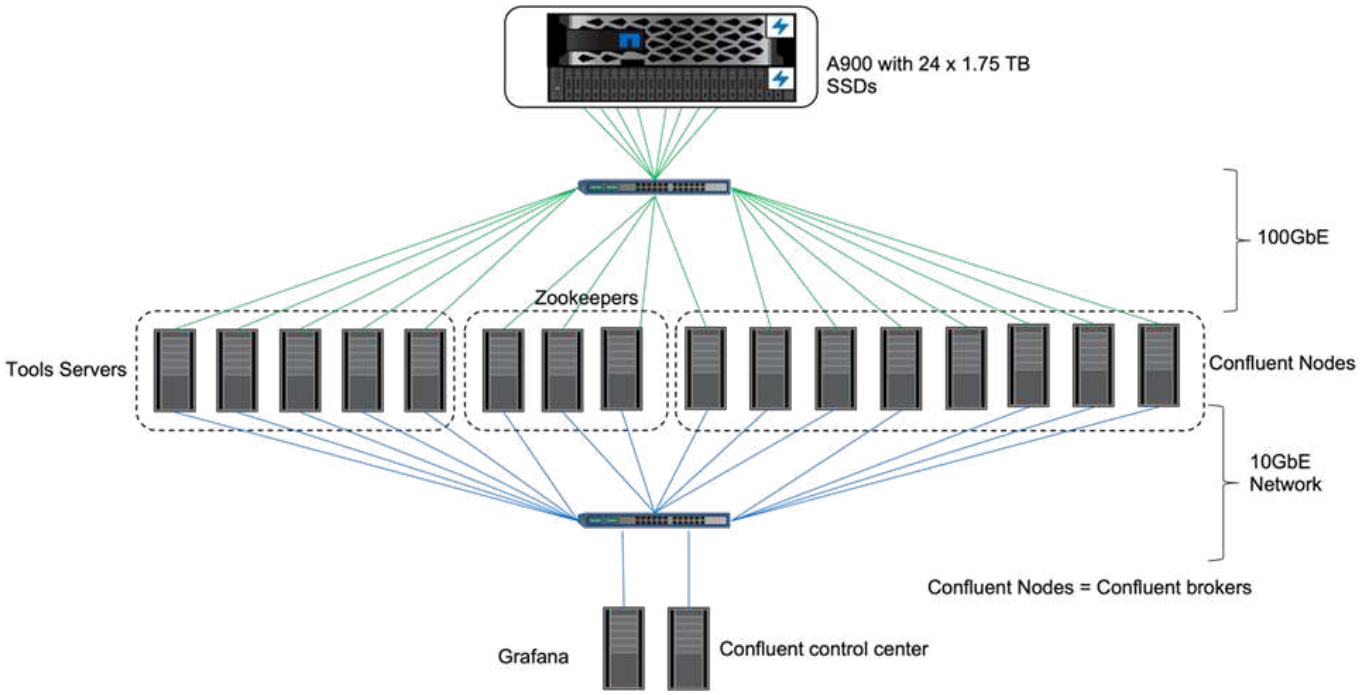
融合性能验证

我们已使用Confluent Platform对NetApp ONTAP 上的分层存储执行了验证。NetApp和Confluent团队共同执行了此验证、并运行了所需的测试用例。

设置冲突

在设置中、我们使用了三个Zookepers、五个代理和五个测试服务器、这些服务器具有256 GB RAM和16个CPU。对于NetApp存储、我们将ONTAP 与AFF A900 HA对结合使用。存储和代理通过100GbE连接进行连接。

下图显示了用于分层存储验证的配置的网络拓扑。



工具服务器充当向Confluent节点发送或从Confluent节点接收事件的应用程序客户端。

融合的分层存储配置

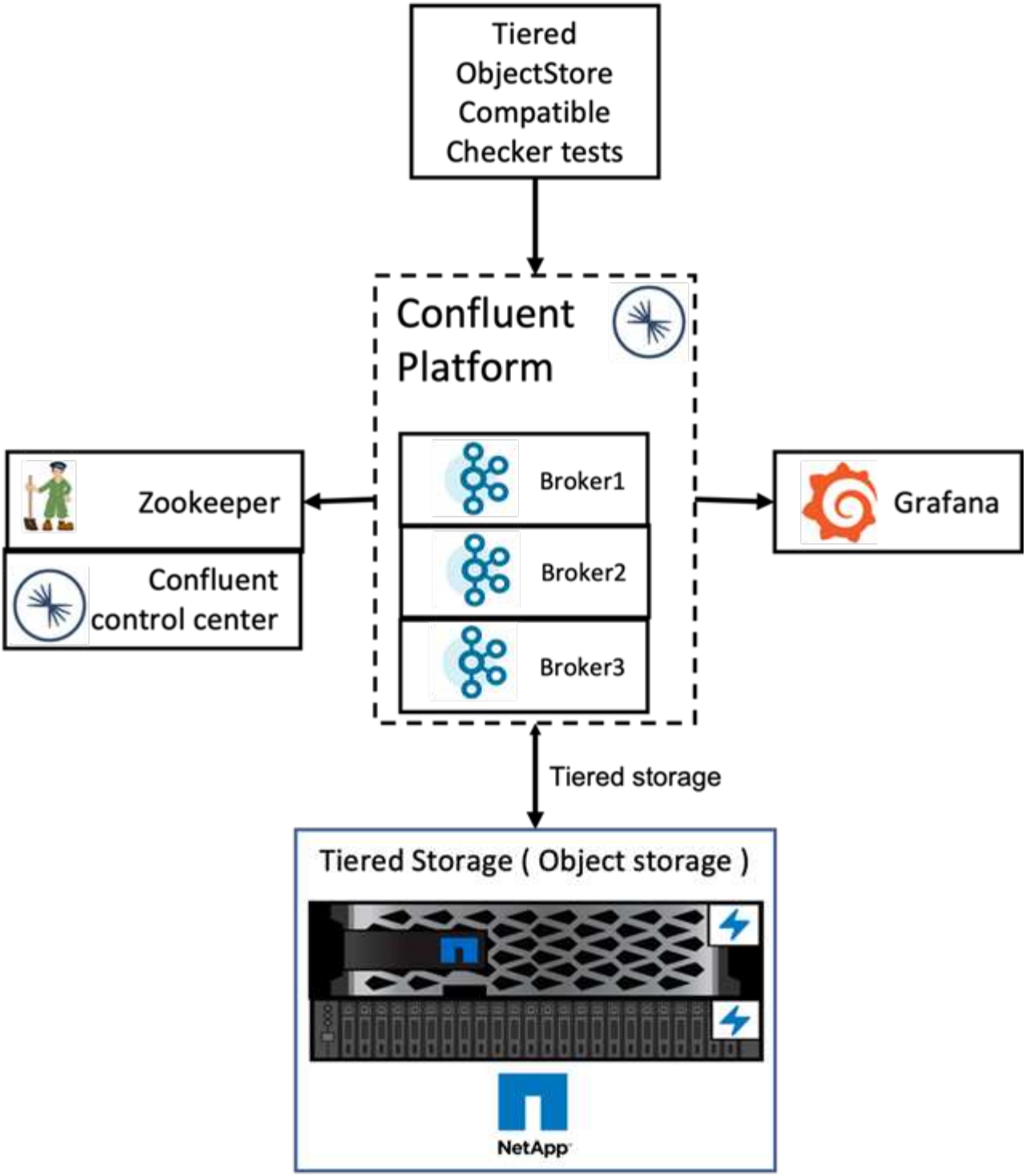
我们使用了以下测试参数：

```
confluent.tier.fetcher.num.threads=80
confluent.tier.archiver.num.threads=80
confluent.tier.enable=true
confluent.tier.feature=true
confluent.tier.backend=S3
confluent.tier.s3.bucket=kafkabucket1-1
confluent.tier.s3.region=us-east-1
confluent.tier.s3.cred.file.path=/data/kafka/.ssh/credentials
confluent.tier.s3.aws.endpoint.override=http://wle-mendocino-07-08/
confluent.tier.s3.force.path.style.access=true
bootstrap.server=192.168.150.172:9092,192.168.150.120:9092,192.168.150.164
:9092,192.168.150.198:9092,192.168.150.109:9092,192.168.150.165:9092,192.1
68.150.119:9092,192.168.150.133:9092
debug=true
jmx.port=7203
num.partitions=80
num.records=200000000
#object PUT size - 512MB and fetch 100MB – netapp
segment.bytes=536870912
max.partition.fetch.bytes=1048576000
#GET size is max.partition.fetch.bytes/num.partitions
length.key.value=2048
trogdor.agent.nodes=node0,node1,node2,node3,node4
trogdor.coordinator.hostname.port=192.168.150.155:8889
num.producers=20
num.head.consumers=20
num.tail.consumers=1
test.binary.task.max.heap.size=32G
test.binary.task.timeout.sec=3600
producer.timeout.sec=3600
consumer.timeout.sec=3600
```

为了进行验证、我们将ONTAP 与HTTP协议结合使用、但HTTPS也可以正常工作。访问密钥和机密密钥存储在 `confuent.tier.s3.cred.file.path` 参数中提供的文件名中。

NetApp存储控制器—ONTAP

我们在ONTAP 中配置了一个HA对配置以进行验证。



验证结果

我们已完成以下五个测试案例以进行验证。前两项是功能测试，其余三项是性能测试。

对象存储正确性测试

此测试使用API调用对用于分层存储的对象存储执行基本操作、例如GET、PUT和DELETE。

分层功能正确性测试

此测试将检查对象存储的端到端功能。它会创建一个主题，为新创建的主题生成一个事件流，等待代理将这些分段归档到对象存储，使用事件流，并验证已用流与已生成流的匹配情况。我们执行此测试时，无论是否注入了对象存储故障。我们通过在ONTAP 中的一个节点中停止服务管理器服务并验证端到端功能是否适用于对象存储来模拟节点故障。

层提取基准测试

此测试验证了分层对象存储的读取性能，并检查了基准测试生成的区块在负载过重时的范围提取读取请求。在此基准测试中， Confluent 开发了自定义客户端来处理层提取请求。

生成并使用工作负载生成器

此测试会通过归档区块间接在对象存储上生成写入工作负载。读取工作负载（区块读取）是在使用者组提取区块时从对象存储生成的。此工作负载由TOCC脚本生成。此测试检查了并行线程中对象存储上的读写性能。与分层功能正确性测试一样，我们测试了是否存在对象存储故障注入。

保留工作负载生成器

此测试检查了在主题保留工作负载繁重的情况下对象存储的删除性能。保留工作负载是使用TOCC脚本生成的、该脚本会与测试主题并行生成许多消息。本测试主题使用主动式基于大小和基于时间的保留设置进行配置，此设置会导致从对象存储中持续清除事件流。然后，这些区块会归档。这导致代理在对象存储中删除了许多内容、并收集了对象存储删除操作的性能。

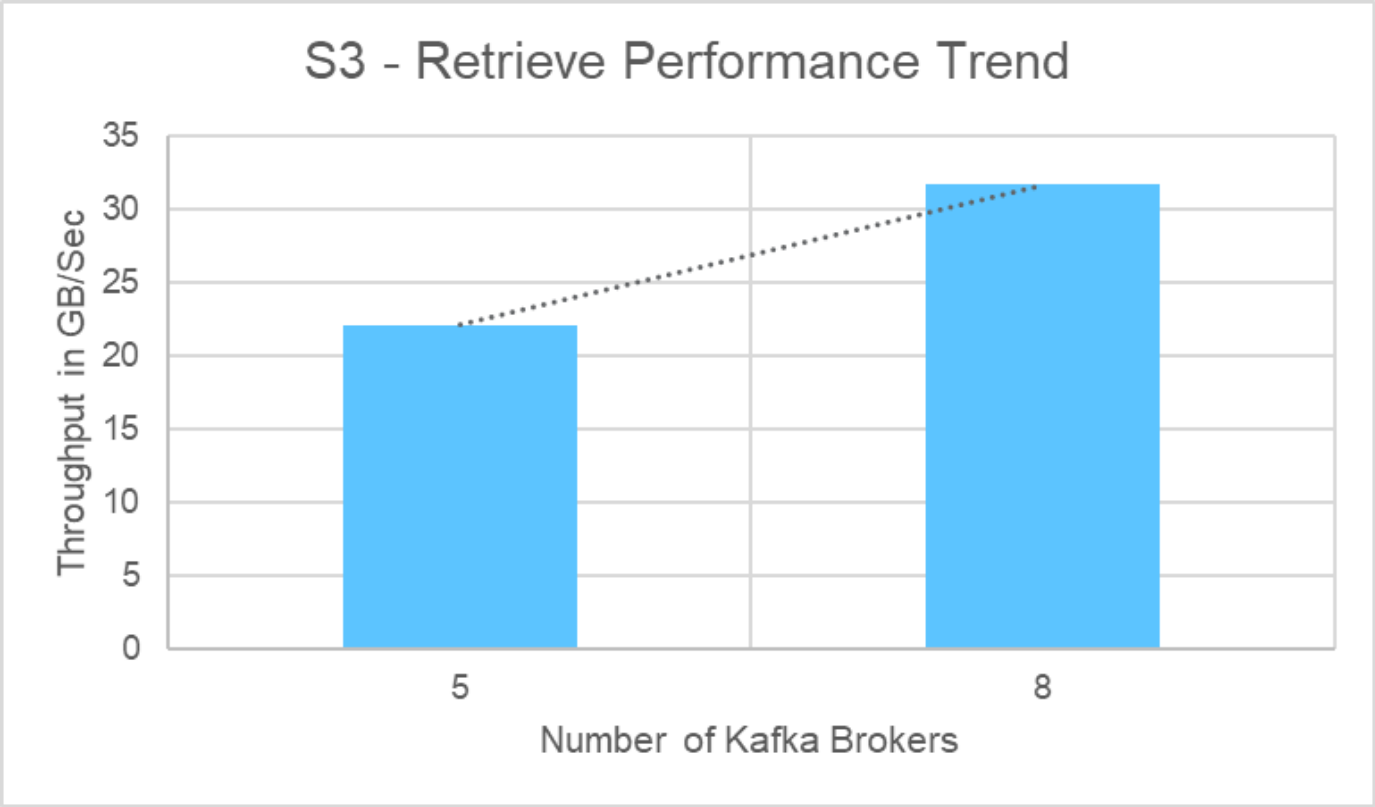
有关验证详细信息、请参见 "两者结合" 网站。

使用生产用工作负载生成器进行性能测试

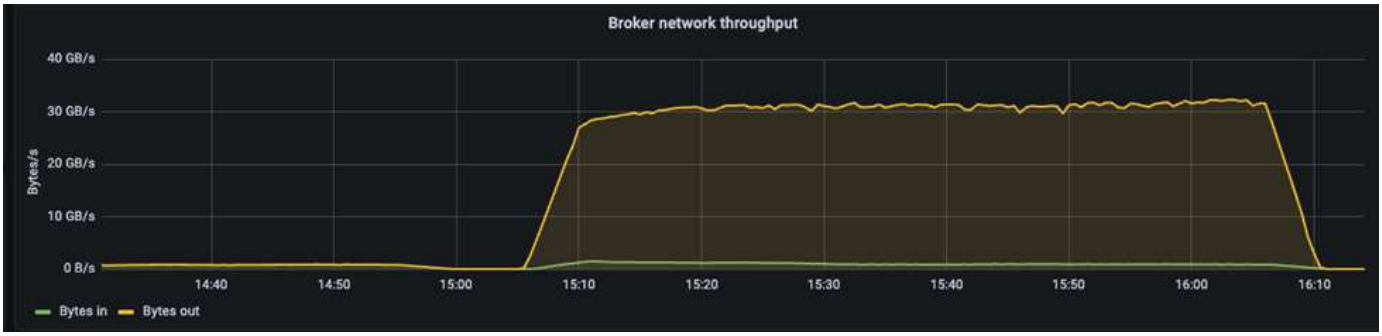
在使用一个AFF A900 HA对NetApp存储控制器的生产用工作负载期间、我们使用五个或八个代理节点执行分层存储测试。根据我们的测试结果、完成时间和性能结果会随着代理节点的数量进行扩展、直到AFF A900资源利用率达到百分之一百为止。ONTAP 存储控制器设置至少需要一个HA对。

根据Confluent代理节点的数量、S3检索操作的性能呈线性增长。ONTAP 存储控制器在一个部署中最多支持12个HA对。

下图显示了具有五个或八个代理节点的S3分层流量组合。我们最大限度地提高了AFF A900单HA对的性能。



下图显示了Kafka吞吐量约为31.74 GBps。



我们还在ONTAP 存储控制器`perfstat`报告中观察到类似的吞吐量。

```
object_store_server:wle-mendocino-07-08:get_data:34080805907b/ s
object_store_server:wle-mendocino-07-08:put_data:484236974b/ s
```

性能最佳实践准则

此页面介绍了在此解决方案 中提高性能的最佳实践。

- 对于ONTAP 、如果可能、请使用GET大小>=1MB。
- 通过在代理节点上的`server.properties`中增加`num.network.threads`和`num.io.threads`、您可以将更多的分层活动推送到S3层。这些结果会在`num.network.threads`和`num.io.threads`设置为32的情况下显示。

- S3存储分段应针对每个成员聚合的八个成分卷。
- 驱动S3流量的以太网链路应尽可能在存储和客户端上使用9k的MTU。

结论

此验证测试在使用NetApp ONTAP 存储控制器的Confluent上达到了31.74 GBps的分层吞吐量。

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，请查看以下文档和 / 或网站：

- 什么是Confluent?

  "https://www.confluent.io/apache-kafka-vs-confluent/"

- S3-sink 参数详细信息

  "https://docs.confluent.io/kafka-connect-s3-sink/current/configuration_options.html#s3-configuration-options"

- Apache Kafka

  "https://en.wikipedia.org/wiki/Apache_Kafka"

- ONTAP 最佳实践中的S3

  https://www.netapp.com/pdf.html?item=/media/17219-tr4814.pdf

- S3对象存储管理

  "https://docs.netapp.com/us-en/ontap/s3-config/s3-support-concept.html"

- NetApp 产品文档

  "https://www.netapp.com/support-and-training/documentation/"

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。