NetApp StorageGRID与 Splunk SmartStore

NetApp artificial intelligence solutions

NetApp
December 04, 2025

This PDF was generated from https://docs.netapp.com/zh-cn/netapp-solutions-ai/data-analytics/stgr-splunkss-introduction.html on December 04, 2025. Always check docs.netapp.com for the latest.

目录

NetApp StorageGRID与 Splunk SmartStore . . . 1
TR-4869： NetApp StorageGRID与 Splunk SmartStore . . . 1
概述 . . . 1
关于NetApp StorageGRID . . . 1
关于Splunk Enterprise . . . 3
关于Splunk SmartStore . . . 3
解决方案概述 . . . 3
NetAppStorageGRID . . . 3
Splunk Enterprise . . . 3
Splunk SmartStore . . . 4
此解决方案的优势 . . . 4
Splunk 架构 . . . 4
关键定义 . . . 4
Splunk 分布式部署 . . . 5
Splunk SmartStore . . . 6
Splunk SmartStore 数据流 . . . 7
软件要求 . . . 8
单站点和多站点要求 . . . 8
硬件要求 . . . 10
Splunk 设计 . . . 12
适用于 Splunk SmartStore 的灵活StorageGRID功能 . . . 14
使用网格管理器进行简单管理 . . . 14
适用于 Splunk 的NetApp StorageGRID应用程序 . . . 15
ILM 策略 . . . 15
性能 . . . 15
负载均衡器和端点配置 . . . 15
智能分层和成本节约 . . . 16
单站点 SmartStore 性能 . . . 16
配置 . . . 19
SmartStore远程商店性能验证 . . . 19
StorageGRID性能 . . . 24
StorageGRID硬件使用情况 . . . 25
采用NetApp存储控制器的 SmartStore - 为客户带来好处 . . . 26
结束语 . . . 27
在哪里可以找到更多信息 . . . 27

NetApp StorageGRID与 Splunk SmartStore

TR-4869： NetApp StorageGRID与 Splunk SmartStore

Splunk Enterprise 是市场领先的安全信息和事件管理 (SIEM) 解决方案，可推动安全、IT 和 DevOps 团队取得成果。

概述

数据量继续以指数级增长，为能够利用这一巨大资源的企业创造了巨大的机会。 Splunk Enterprise 在更广泛的使用案例中得到广泛应用。随着用例的增长，Splunk Enterprise 提取和处理的数据量也在增加。 Splunk Enterprise 的传统架构是分布式横向扩展设计，提供出色的数据访问和可用性。然而，使用这种架构的企业面临着与扩展相关的成本不断增长的问题，以满足快速增长的数据量。

采用NetApp StorageGRID的 Splunk SmartStore 通过提供一种计算和存储分离的新部署模型解决了这一难题。该解决方案还允许客户跨单个和多个站点进行扩展，从而为 Splunk Enterprise 环境提供无与伦比的规模和弹性，同时通过允许计算和存储独立扩展并为经济高效的基于云的 S3 对象存储添加智能分层来降低成本。

该解决方案在保持搜索性能的同时优化了本地存储的数据量，允许按需扩展计算和存储。 SmartStore 自动评估数据访问模式，以确定哪些数据需要进行实时分析，哪些数据应该驻留在成本较低的 S3 对象存储中。

本技术报告概述了NetApp为 Splunk SmartStore 解决方案带来的优势，同时演示了在您的环境中设计和调整 Splunk SmartStore 大小的框架。最终结果是一个简单、可扩展且有弹性的解决方案，可提供极具吸引力的 TCO。 StorageGRID提供可扩展且经济高效的基于 S3 协议/API 的对象存储（也称为远程存储），使组织能够以较低的成本扩展其 Splunk 解决方案，同时提高弹性。

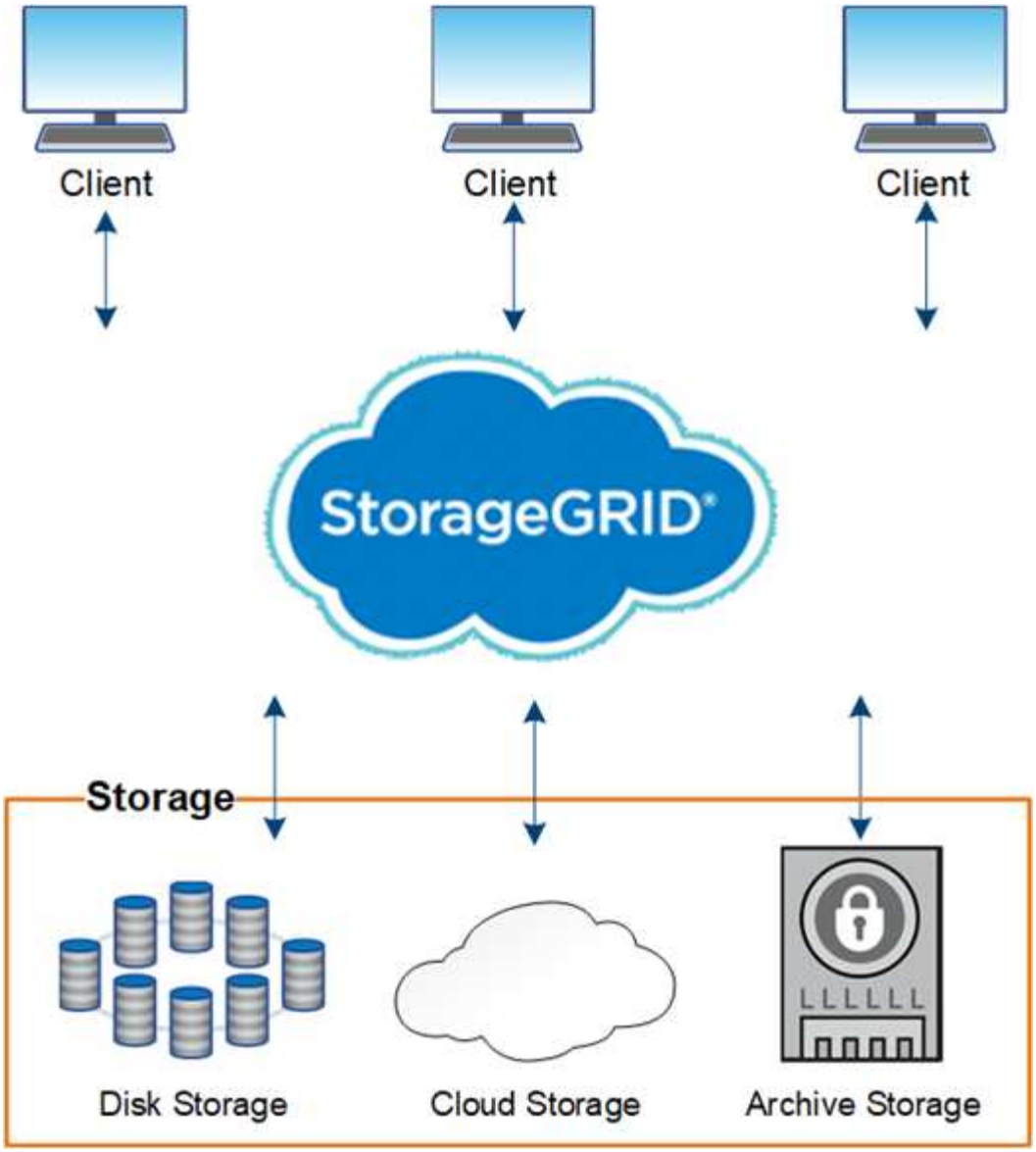
Splunk SmartStore 将对象存储称为远程存储或远程存储层。

关于NetApp StorageGRID

NetApp StorageGRID是一种软件定义的对象存储解决方案，适用于大型档案、媒体存储库和 Web 数据存储。借助StorageGRID， NetApp利用二十年来提供业界领先的创新和数据管理解决方案的经验，同时帮助组织管理和最大化其内部以及公共、私有或混合云部署中的信息价值。

StorageGRID为大规模非结构化数据提供安全、持久的存储。集成的、元数据驱动的生命周期管理策略可优化数据在整个生命周期中的存储位置。将内容放置在正确的位置、正确的时间以及正确的存储层以降低成本。单一命名空间允许通过单一调用访问数据，而不管StorageGRID存储的地理位置如何。客户可以在数据中心之间和云基础设施中部署和管理多个StorageGRID实例。

StorageGRID系统由全球分布、冗余、异构的节点组成，可以与现有和下一代客户端应用程序集成。



IDC MarketScape 最近在最新报告《IDC MarketScape：2019 年全球基于对象的存储供应商评估》中将NetApp评为领导者。 StorageGRID拥有近 20 年在要求最严格的行业中进行生产部署的经验，是非结构化数据领域公认的领导者。

借助StorageGRID，您可以实现以下目标：

- 部署多个StorageGRID实例，通过可轻松扩展到数百 PB 的单个命名空间访问数据中心和云之间任何位置的数据。
- 提供跨基础设施部署和集中管理的灵活性。
- 利用分层擦除编码 (EC) 提供无与伦比的耐用性，耐用性达到 15 个 9。
- 通过与 Amazon S3 Glacier 和 Azure Blob 进行验证的集成，实现更多混合多云功能。
- 通过防篡改数据保留满足监管义务并促进合规性，无需专有 API 或供应商锁定。

有关StorageGRID如何帮助您解决最复杂的非结构化数据管理问题的更多信息，请参阅 "NetApp StorageGRID主页"。

关于Splunk Enterprise

Splunk Enterprise 是一个将数据转化为行动的平台。日志文件、网站、设备、传感器和应用程序等各种来源生成的数据被发送到 Splunk Indexer 并由其解析，从而使您可以从数据中获得丰富的见解。它可能识别数据泄露、指出客户和产品趋势、寻找优化基础设施的机会或在各种用例中创建可操作的见解。

关于Splunk SmartStore

Splunk SmartStore 扩展了 Splunk 架构的优势，同时简化了其经济高效扩展的能力。计算和存储资源的分离导致索引器节点针对 I/O 进行了优化，并且存储需求显著减少，因为它们仅将一部分数据存储为缓存。当只需要其中一种资源时，您不必添加额外的计算或存储，从而可以实现显著的成本节约。您可以使用经济高效且易于扩展的基于 S3 的对象存储，这进一步简化了环境、降低了成本并允许您维护更庞大的数据集。

Splunk SmartStore 为组织带来巨大价值，包括：

- 通过将热数据移动到成本优化的 S3 对象存储来降低存储成本
- 通过分离存储和计算实现无缝扩展
- 利用弹性云原生存储简化业务连续性

解决方案概述

本页介绍了完成此解决方案所使用的组件，包括NetApp StorageGRID、Splunk Enterprise 和 Splunk SmartStore。

NetAppStorageGRID

NetApp StorageGRID是一个高性能且经济高效的对象存储平台。它使用分布式、基于节点的网格架构提供智能、策略驱动的全球数据管理。它通过其无处不在的全局对象命名空间与复杂的数据管理功能相结合，简化了 PB 级非结构化数据和数十亿个对象的管理。单次调用对象访问可跨站点扩展，并简化高可用性架构，同时确保无论站点或基础设施是否中断，都能持续进行对象访问。

多租户允许多个云和企业非结构化数据应用程序在同一网格内得到安全地服务，从而增加了StorageGRID的投资回报率和用例。可以使用元数据驱动的对象生命周期策略创建多个服务级别，从而优化跨多个地区的耐用性、保护性、性能和局部性。随着需求的变化，用户可以调整策略并且无中断地重新调整数据格局。

SmartStore 利用StorageGRID作为远程存储层，并允许客户部署多个地理分布的站点，以实现强大的可用性和耐用性，并以单个对象命名空间的形式呈现。这使得 Splunk SmartStore 能够利用StorageGRID的高性能、高密度容量以及使用单个 URL 与对象交互扩展到多个物理站点上的数百个节点的能力。此单一 URL 还允许在不中断的情况下进行存储扩展、升级和修复，甚至超越单个站点。 StorageGRID独特的数据管理策略引擎提供了优化的性能和耐用性水平，并符合数据局部性要求。

Splunk Enterprise

Splunk 是机器生成数据收集和分析领域的领导者，通过其运营分析功能帮助简化和现代化 IT。它还扩展到商业分析、安全和物联网用例。存储是成功部署 Splunk 软件的关键因素。

机器生成的数据是增长最快的大数据类型。其格式难以预测，并且来自许多不同的来源，通常速率很高且数量巨大。这些工作负载特征通常被称为数字排气。 Splunk SmartStore 有助于理解这些数据并提供智能数据分层，以便在最具成本效益的存储层上优化放置热数据和温数据。

Splunk SmartStore

Splunk SmartStore 是一种索引器功能，它使用对象存储（也称为远程存储或远程存储层）例如StorageGRID通过 S3 协议存储热数据。

随着部署的数据量增加，对存储的需求通常会超过对计算机资源的需求。 SmartStore 允许您通过分别扩展计算和存储来经济高效地管理索引器存储和计算资源。

SmartStore 引入了使用 S3 协议的远程存储层和缓存管理器。这些功能允许数据驻留在本地索引器或远程存储上。缓存管理器位于索引器上，负责管理索引器和远程存储层之间的数据移动。数据与存储桶元数据一起存储在存储桶（热和温）中。

使用 SmartStore，您可以将索引器存储占用空间降至最低，并选择 I/O 优化的计算资源，因为大多数数据都驻留在远程存储层上。索引器维护一个本地缓存，代表返回请求和预测结果所需的最少数据量。本地缓存包含热存储桶、参与活动或最近搜索的热存储桶副本以及存储桶元数据。

带有StorageGRID的 Splunk SmartStore 使客户能够通过高性能且经济高效的远程存储逐步扩展环境，同时为整体解决方案提供高度的弹性。这使得客户可以在任何给定时间添加任何给定数量的任何组件（热存储和/或温 S3 存储），无论他们是需要更多索引器、更改数据保留，还是在不造成任何中断的情况下增加摄取率。

此解决方案的优势

该解决方案允许添加计算、热存储或 S3 资源，以满足单站点和多站点部署中用户数量或摄取率不断增长的需求。

- 表现。 Splunk SmartStore 和NetApp StorageGRID的结合使用对象存储在热存储桶和温存储桶之间实现数据的快速迁移。 StorageGRID通过为大型对象工作负载提供快速性能来加速迁移过程。
- 多站点就绪。 StorageGRID分布式架构允许 Splunk SmartStore 通过单个全局命名空间扩展跨单个和多个站点的部署，无论数据位于何处，都可以从任何站点访问数据。
- *提高了可扩展性。*独立于计算资源扩展存储资源，以满足 Splunk 环境中不断变化的需求，从而提供更好的 TCO。
- *容量。*使用StorageGRID将单个命名空间扩展到 560PB 以上，满足 Splunk 部署中快速增长的容量。
- *数据可用性。*使用元数据驱动的策略来优化数据可用性、性能、地理分布、保留、保护和存储成本，这些策略可以随着数据的业务价值的发展而动态调整。

使用 SmartStore 缓存提高性能，它是索引器的一个组件，用于处理本地（热）和远程（温）存储之间的存储桶副本传输。此解决方案的 Splunk 规模基于 "Splunk 提供的指南"。该解决方案允许添加计算、热存储或 S3 资源，以满足单站点和多站点部署中用户数量或摄取率不断增长的需求。

Splunk 架构

本节介绍 Splunk 架构，包括关键定义、Splunk 分布式部署、Splunk SmartStore、数据流、硬件和软件要求、单站点和多站点要求等。

关键定义

接下来的两个表列出了分布式 Splunk 部署中使用的 Splunk 和NetApp组件。

此表列出了分布式 Splunk Enterprise 配置的 Splunk 硬件组件。

Splunk 组件	任务
索引器	Splunk Enterprise 数据存储库
通用转发器	负责提取数据并将数据转发给索引器
搜索头	用于在索引器中搜索数据的用户前端
集群主节点	管理索引器和搜索头的 Splunk 安装
监控控制台	整个部署中使用的集中监控工具
许可证主控	许可证管理员处理 Splunk Enterprise 许可
部署服务器	更新配置并将应用程序分发到处理组件
存储组件	任务
NetApp AFF	用于管理热层数据的全闪存存储。也称为本地存储。
NetAppStorageGRID	用于管理热层数据的 S3 对象存储。 SmartStore 使用它在热层和温层之间移动数据。也称为远程存储。

下表列出了 Splunk 存储架构中的组件。

Splunk 组件	任务	负责组件
智能商店	为索引器提供将数据从本地存储分层到对象存储的能力。	Splunk
热的	通用转发器放置新写入数据的着陆点。存储是可写的，数据是可搜索的。该数据层通常由 SSD 或快速 HDD 组成。	ONTAP
缓存管理器	管理索引数据的本地缓存，在搜索时从远程存储中获取热数据，并从缓存中逐出最不常用的数据。	智能商店
温暖的	数据按逻辑滚动到存储桶，首先从热层重命名为暖层。此层内的数据受到保护，并且与热层一样，可以由更大容量的 SSD 或 HDD 组成。使用常见的数据保护解决方案支持增量备份和完整备份。	StorageGRID

Splunk 分布式部署

为了支持数据来自多台机器的更大环境，您需要处理大量数据。如果许多用户需要搜索数据，您可以通过在多台机器上分发 Splunk Enterprise 实例来扩展部署。这被称为分布式部署。

在典型的分布式部署中，每个 Splunk Enterprise 实例执行一项专门的任务，并驻留在与主要处理功能相对应的三个处理层之一上。

下表列出了 Splunk Enterprise 处理层。

层级	组件	描述
数据输入	货运代理	转发器消费数据，然后将数据转发给一组索引器。
索引	索引器	索引器对通常从一组转发器接收的传入数据进行索引。索引器将数据转换为事件并将事件存储在索引中。索引器还根据搜索头的搜索请求搜索索引数据。
搜索管理	搜索头	搜索头是搜索的中心资源。集群中的搜索头是可互换的，并且可以从搜索头集群的任何成员访问相同的搜索、仪表板、知识对象等。

下表列出了分布式 Splunk Enterprise 环境中使用的重要组件。

组件	描述	责任
索引集群主节点	协调索引器集群的活动和更新	索引管理
索引集群	配置为相互复制数据的 Splunk Enterprise 索引器组	索引
搜索头部署器	处理集群主控的部署和更新	搜索头管理
搜索头集群	一组搜索头，作为搜索的中心资源	搜索管理
负载均衡器	由集群组件使用，以处理搜索头、索引器和 S3 目标不断增长的需求，从而在集群组件之间分配负载。	集群组件的负载管理

了解 Splunk Enterprise 分布式部署的以下优势：

- 访问多样化或分散的数据源
- 提供处理任何规模和复杂程度的企业数据需求的功能
- 通过数据复制和多站点部署实现高可用性并确保灾难恢复

Splunk SmartStore

SmartStore 是一种索引器功能，它使远程对象存储（如 Amazon S3）能够存储索引数据。随着部署的数据量增加，对存储的需求通常会超过对计算资源的需求。 SmartStore 允许您通过单独扩展这些资源来经济高效地管理索引器存储和计算资源。

SmartStore 引入了远程存储层和缓存管理器。这些功能允许数据驻留在本地索引器上或远程存储层上。缓存管理器管理索引器和在索引器上配置的远程存储层之间的数据移动。

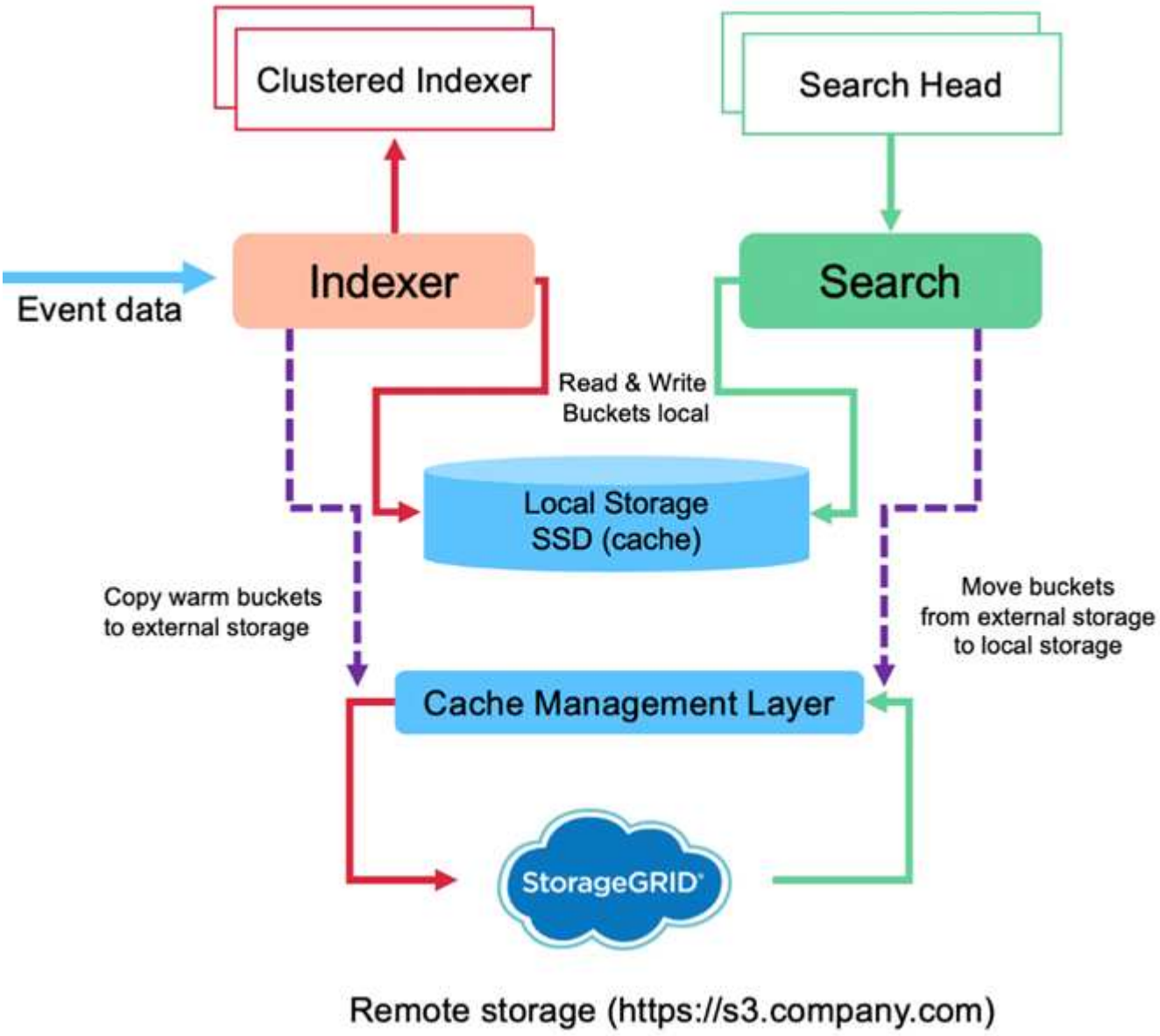
使用 SmartStore，您可以将索引器存储占用空间降至最低，并选择针对 I/O 优化的计算资源。大多数数据驻留在远程存储上。索引器维护一个包含最少量数据的本地缓存：热存储桶、参与活动或最近搜索的热存储桶副本以及存储桶元数据。

Splunk SmartStore 数据流

当来自各个来源的数据到达索引器时，数据会被索引并本地保存在热存储桶中。索引器还将热存储桶数据复制到目标索引器。到目前为止，数据流与非 SmartStore 索引的数据流相同。

当热桶变暖时，数据流就会分叉。源索引器将热存储桶复制到远程对象存储（远程存储层），同时将现有副本保留在其缓存中，因为搜索往往会遇到最近索引的数据。但是，目标索引器会删除其副本，因为远程存储无需维护多个本地副本即可提供高可用性。存储桶的主副本现在位于远程存储中。

下图显示了 Splunk SmartStore 数据流。



索引器上的缓存管理器是 SmartStore 数据流的核心。它根据需要从远程存储中获取存储桶的副本来处理搜索请求。它还会从缓存中逐出较旧或搜索较少的存储桶副本，因为它们参与搜索的可能性会随着时间的推移而降低。

缓存管理器的工作是优化可用缓存的使用，同时确保搜索可以立即访问所需的存储桶。

软件要求

下表列出了实施该解决方案所需的软件组件。解决方案实施过程中所使用的软件组件可能会根据客户要求而有所不同。

产品系列	产品名称	产品版本	操作系统
NetAppStorageGRID	StorageGRID对象存储	11.6	不适用
CentOS	CentOS	8.1	CentOS 7.x
Splunk Enterprise	Splunk Enterprise 与 SmartStore	8.0.3	CentOS 7.x

单站点和多站点要求

在企业 Splunk 环境（中型和大型部署）中，数据源自多台机器，并且许多用户需要搜索数据，您可以通过在单个和多个站点上分发 Splunk Enterprise 实例来扩展部署。

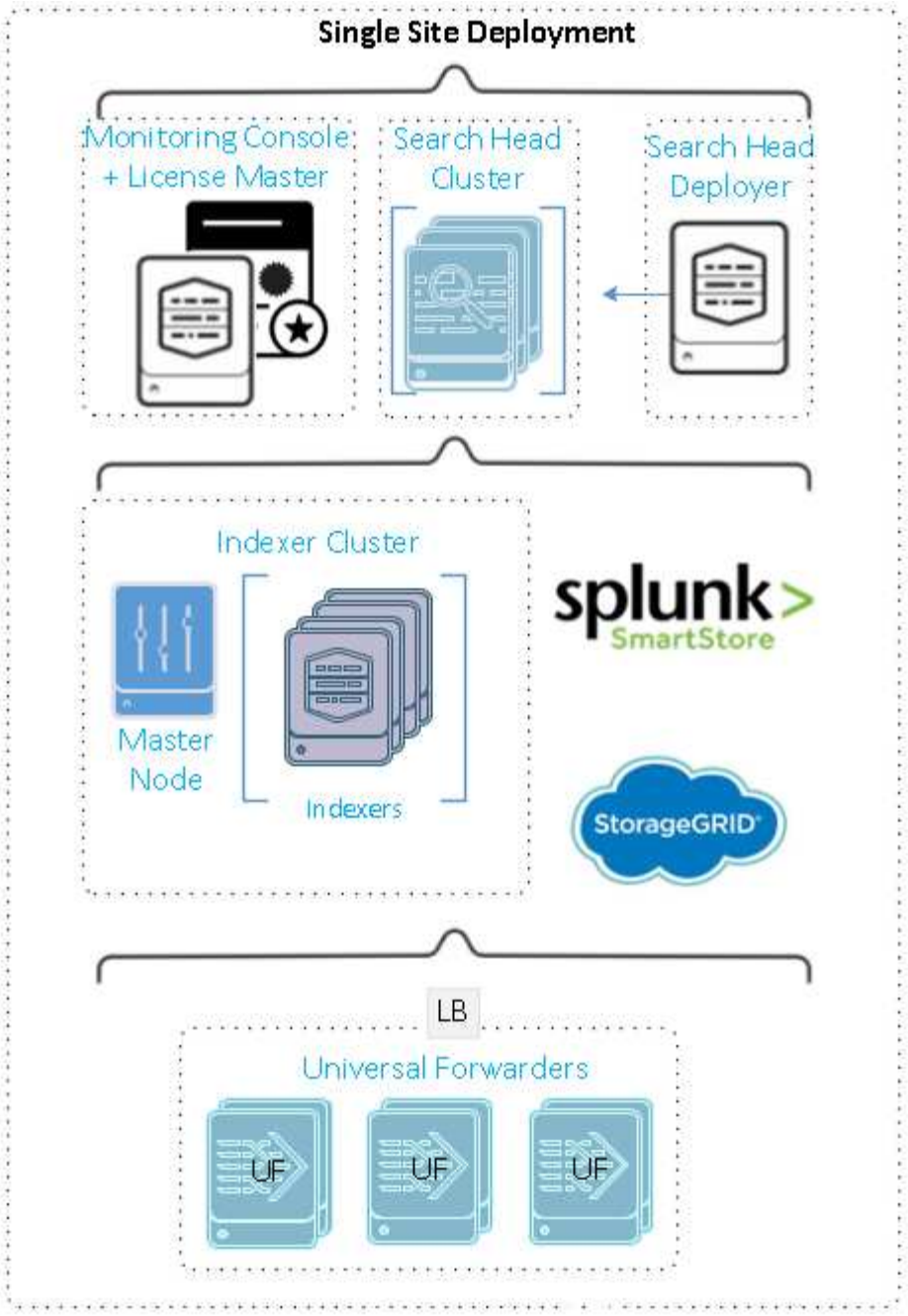
了解 Splunk Enterprise 分布式部署的以下优势：

- 访问多样化或分散的数据源
- 提供处理任何规模和复杂程度的企业数据需求的功能
- 通过数据复制和多站点部署实现高可用性并确保灾难恢复

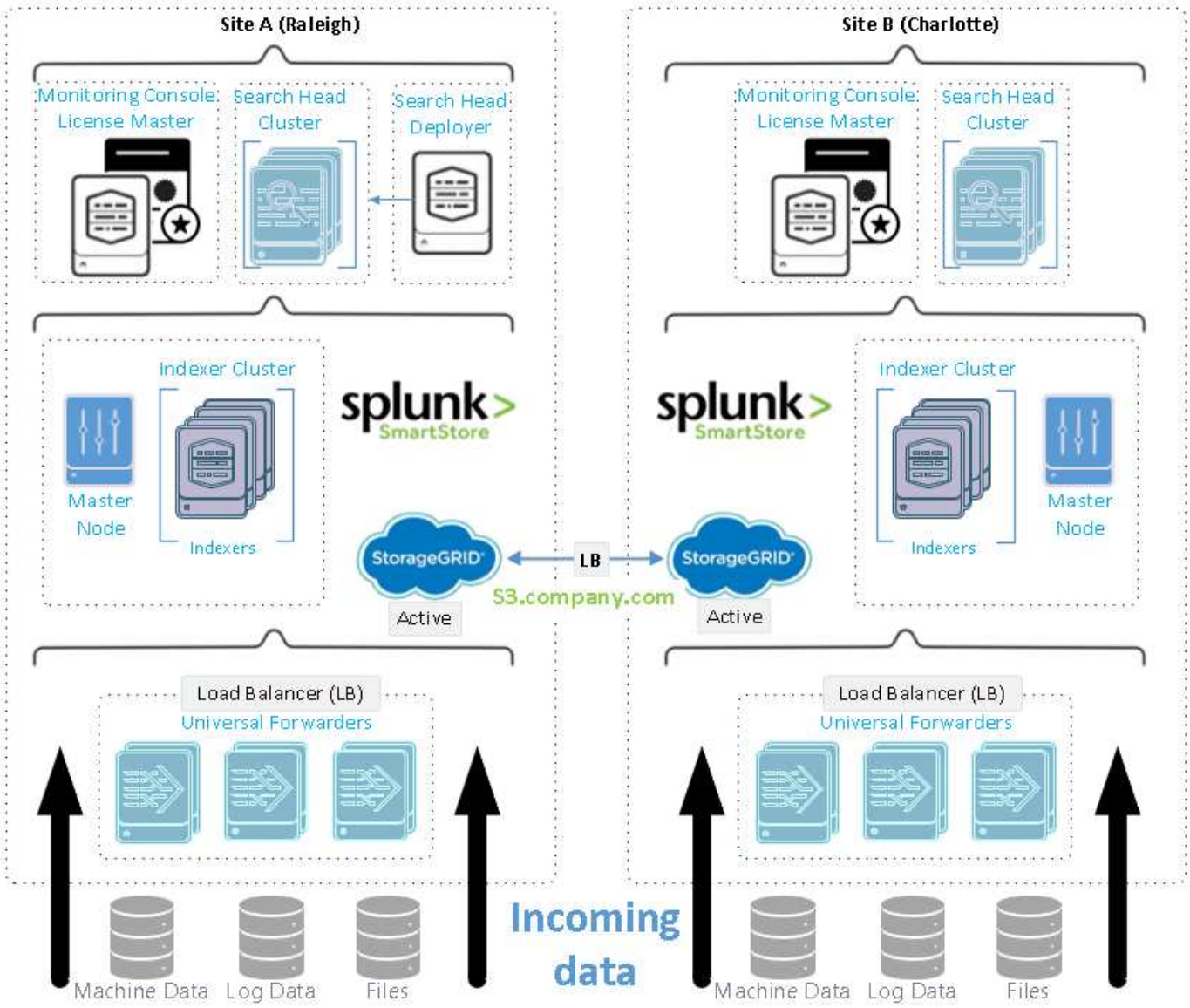
下表列出了分布式 Splunk Enterprise 环境中使用的组件。

组件	描述	责任
索引集群主节点	协调索引器集群的活动和更新	索引管理
索引集群	配置为相互复制数据的 Splunk Enterprise 索引器组	索引
搜索头部署器	处理集群主控的部署和更新	搜索头管理
搜索头集群	一组搜索头，作为搜索的中心资源	搜索管理
负载均衡器	由集群组件使用，以处理搜索头、索引器和 S3 目标不断增长的需求，从而在集群组件之间分配负载。	集群组件的负载管理

该图描绘了单站点分布式部署的示例。



该图描绘了多站点分布式部署的示例。



硬件要求

下表列出了实施该解决方案所需的最少硬件组件数量。解决方案具体实施中使用的硬件组件可能根据客户要求而有所不同。

> 无论您在单个站点还是多个站点部署了 Splunk SmartStore 和StorageGRID ，所有系统都通过StorageGRID GRID Manager 在单一玻璃窗格中进行管理。有关更多详细信息，请参阅“使用网格管理器进行简单管理”部分。

该表列出了单个站点使用的硬件。

硬件	数量	磁盘	可用容量	注
StorageGRID SG1000	1	不适用	不适用	管理节点和负载均衡器
StorageGRID SG6060	4	x48，8TB（NL-SAS 硬盘）	1PB	远程存储

下表列出了用于多站点配置（每个站点）的硬件。

硬件	数量	磁盘	可用容量	注
StorageGRID SG1000	2	不适用	不适用	管理节点和负载均衡器
StorageGRID SG6060	4	x48，8TB（NL-SAS 硬盘）	1PB	远程存储

NetApp StorageGRID负载均衡器：**SG1000**

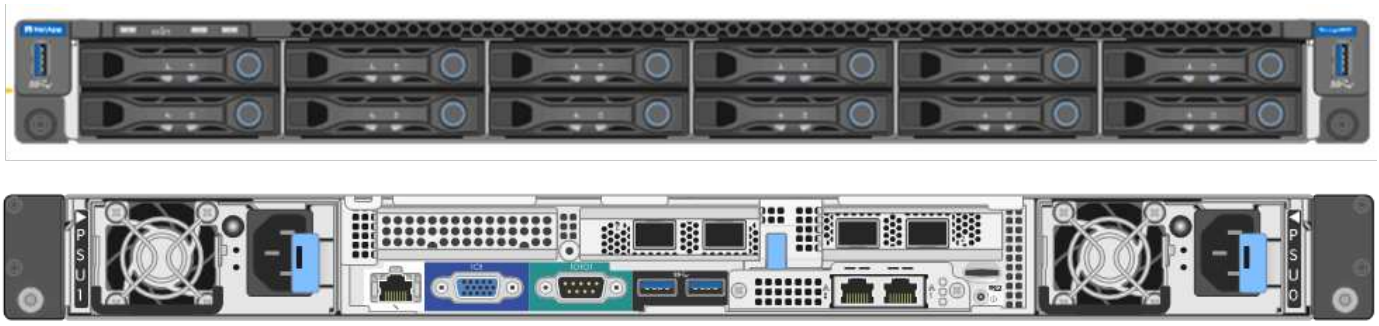
对象存储需要使用负载均衡器来呈现云存储命名空间。 StorageGRID支持来自 F5 和 Citrix 等领先供应商的第三方负载均衡器，但许多客户选择企业级StorageGRID均衡器以实现简单性、弹性和高性能。 StorageGRID负载均衡器可作为虚拟机、容器或专用设备使用。

StorageGRID SG1000 有助于使用高可用性 (HA) 组和 S3 数据路径连接的智能负载平衡。没有其他内部部署对象存储系统提供定制的负载均衡器。

SG1000 设备提供以下功能：

- StorageGRID系统的负载均衡器和管理节点（可选）功能
- StorageGRID Appliance Installer 可简化节点部署和配置
- 简化 S3 端点和 SSL 的配置
- 专用带宽（而不是与其他应用程序共享第三方负载均衡器）
- 高达 4 x 100Gbps 聚合以太网带宽

下图显示了 SG1000 网关服务设备。



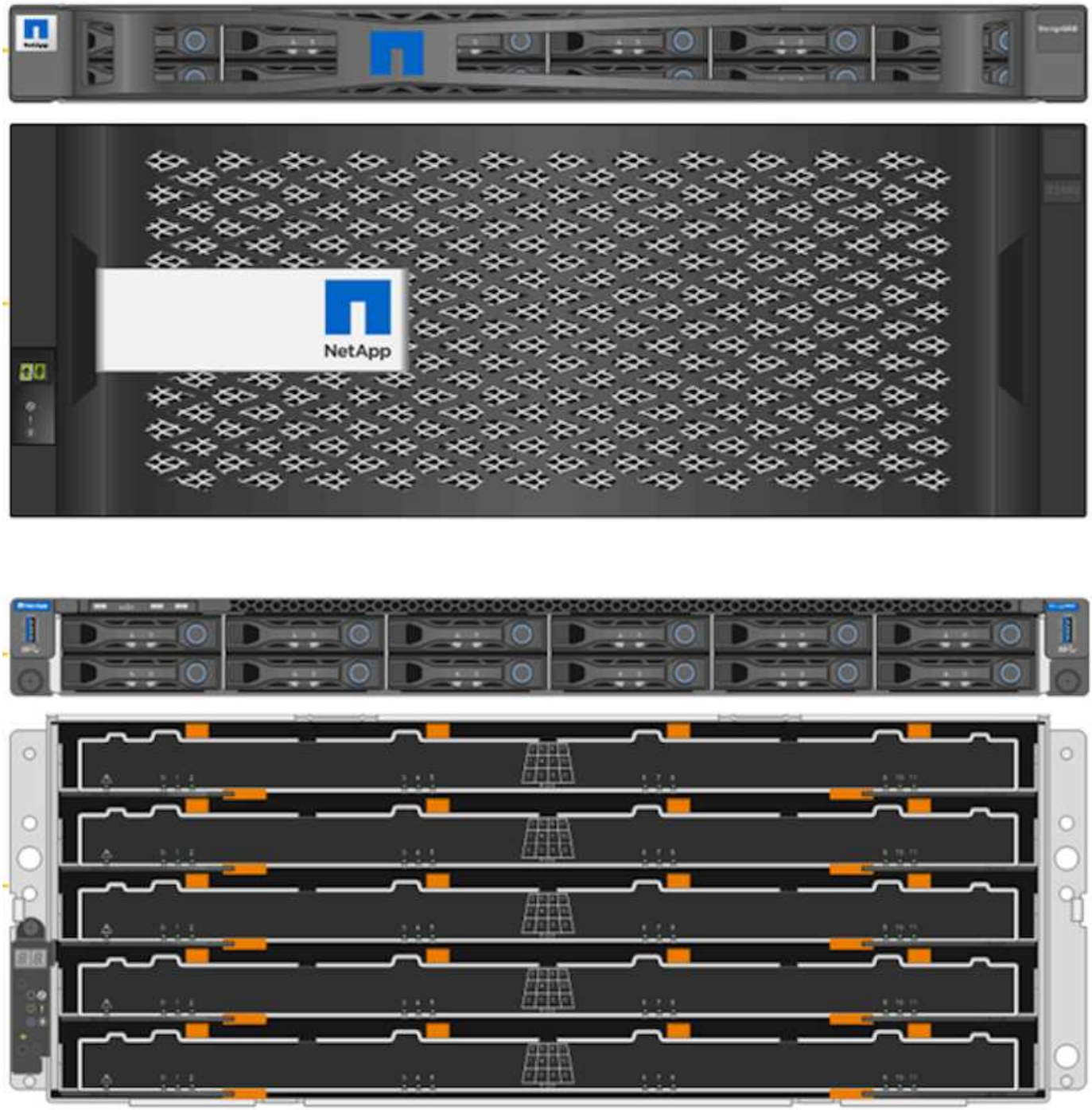
SG6060

StorageGRID SG6060 设备包括一个计算控制器（SG6060）和一个存储控制器架（E 系列 E2860），其中包含两个存储控制器和 60 个驱动器。该设备具有以下功能：

- 在单个命名空间中扩展到 400PB。
- 高达 4x 25Gbps 的聚合以太网带宽。
- 包括StorageGRID Appliance Installer，以简化节点部署和配置。
- 每个 SG6060 设备可以有一个或两个额外的扩展架，总共可容纳 180 个驱动器。

- 两个 E 系列 E2800 控制器（双工配置）提供存储控制器故障转移支持。
- 五抽屉驱动器架，可容纳 60 个 3.5 英寸驱动器（两个固态驱动器和 58 个 NL-SAS 驱动器）。

下图显示了 SG6060 设备。



Splunk 设计

下表列出了单个站点的 Splunk 配置。

Splunk 组件	任务	数量	核心	内存	操作系统
通用转发器	负责提取数据并将数据转发给索引器	4	16 核	32 GB 内存	CentOS 8.1
索引器	管理用户数据	10	16 核	32 GB 内存	CentOS 8.1
搜索头	用户前端在索引器中搜索数据	3	16 核	32 GB 内存	CentOS 8.1
搜索头部署器	处理搜索头集群的更新	1	16 核	32 GB 内存	CentOS 8.1
集群主节点	管理 Splunk 安装和索引器	1	16 核	32 GB 内存	CentOS 8.1
监控控制台和许可证主控器	对整个 Splunk 部署进行集中监控并管理 Splunk 许可证	1	16 核	32 GB 内存	CentOS 8.1

下表描述了多站点配置的 Splunk 配置。

下表列出了多站点配置（站点 A）的 Splunk 配置。

Splunk 组件	任务	数量	核心	内存	操作系统
通用转发器	负责提取数据并将数据转发给索引器。	4	16 核	32 GB 内存	CentOS 8.1
索引器	管理用户数据	10	16 核	32 GB 内存	CentOS 8.1
搜索头	用户前端在索引器中搜索数据	3	16 核	32 GB 内存	CentOS 8.1
搜索头部署器	处理搜索头集群的更新	1	16 核	32 GB 内存	CentOS 8.1
集群主节点	管理 Splunk 安装和索引器	1	16 核	32 GB 内存	CentOS 8.1
监控控制台和许可证主控器	对整个 Splunk 部署进行集中监控并管理 Splunk 许可证。	1	16 核	32 GB 内存	CentOS 8.1

下表列出了多站点配置（站点 B）的 Splunk 配置。

Splunk 组件	任务	数量	核心	内存	操作系统
通用转发器	负责提取数据并将数据转发给索引器	4	16 核	32 GB 内存	CentOS 8.1
索引器	管理用户数据	10	16 核	32 GB 内存	CentOS 8.1

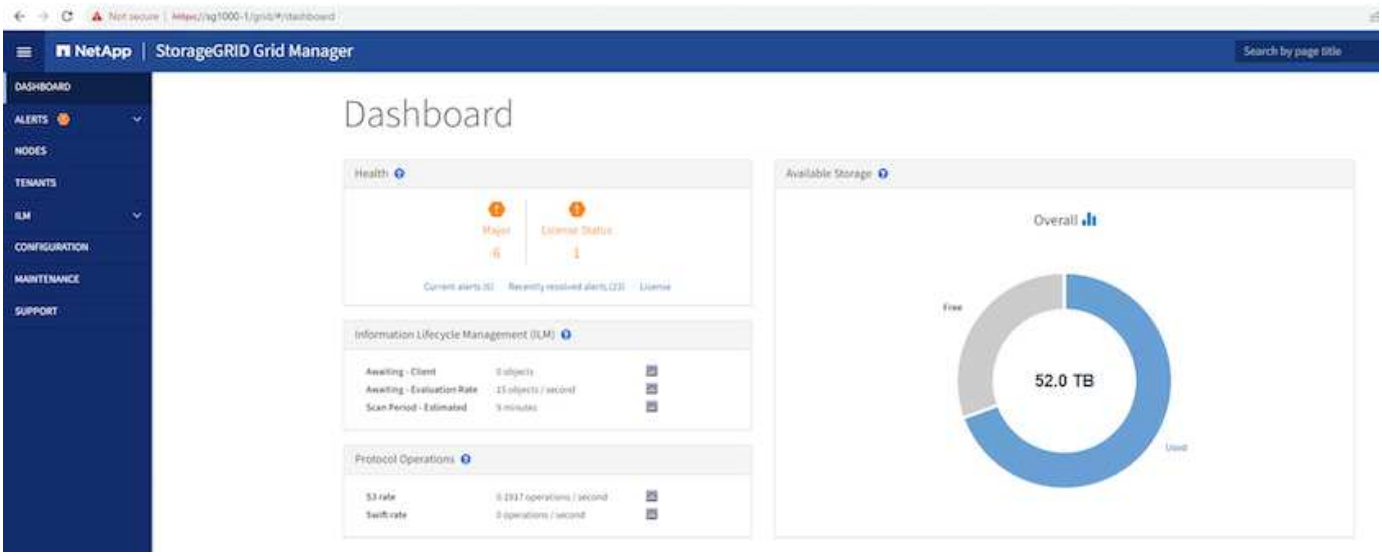
Splunk 组件	任务	数量	核心	内存	操作系统
搜索头	用户前端在索引器中搜索数据	3	16 核	32 GB 内存	CentOS 8.1
集群主节点	管理 Splunk 安装和索引器	1	16 核	32 GB 内存	CentOS 8.1
监控控制台和许可证主控器	对整个 Splunk 部署进行集中监控并管理 Splunk 许可证	1	16 核	32 GB 内存	CentOS 8.1

适用于 Splunk SmartStore 的灵活StorageGRID功能

StorageGRID具有多种功能，用户可以利用这些功能并根据不断变化的环境进行定制。从部署到扩展您的 Splunk SmartStore，您的环境需要快速适应变化，并且不应对 Splunk 造成干扰。 StorageGRID灵活的数据管理策略 (ILM) 和流量分类器 (QoS) 让您可以规划并适应您的环境。

使用网格管理器进行简单管理

Grid Manager 是基于浏览器的图形界面，允许您在单个玻璃窗格中配置、管理和监控全球分布位置的StorageGRID系统，如下图所示。



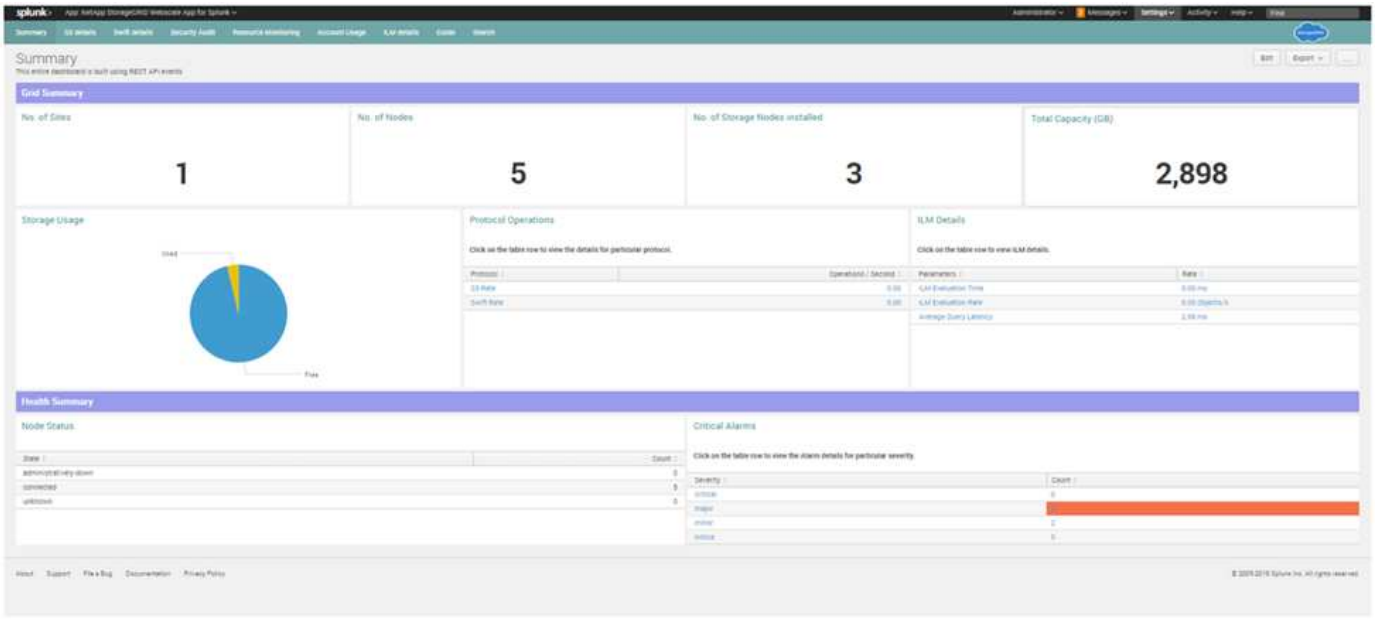
使用网格管理器界面执行以下任务：

- 管理全球分布的 PB 级对象存储库，例如图像、视频和记录。
- 监控网格节点和服务以确保对象可用性。
- 使用信息生命周期管理 (ILM) 规则来管理对象数据随时间推移的放置。这些规则控制着对象数据被摄取后会发生什么、如何防止数据丢失、对象数据存储在何处以及存储多长时间。
- 监控系统内的交易、性能和操作。

适用于 **Splunk** 的**NetApp StorageGRID**应用程序

NetApp StorageGRID App for Splunk 是一款专用于 Splunk Enterprise 的应用程序。此应用程序与 Splunk 的NetApp StorageGRID附加组件配合使用。它提供对StorageGRID健康状况、帐户使用信息、安全审计详细信息、资源使用情况和监控等方面的可见性。

下图显示了适用于 Splunk 的StorageGRID应用程序。



ILM 策略

StorageGRID具有灵活的数据管理策略，包括保留对象的多个副本，并使用 EC（擦除编码）方案（如 2+1 和 4+2（以及许多其他方案））根据特定的性能和数据保护要求存储对象。由于工作负载和需求随时间而变化，ILM 策略通常也必须随时间而变化。修改 ILM 策略是一项核心功能，允许StorageGRID客户快速轻松地适应不断变化的环境。

性能

StorageGRID通过添加更多节点来扩展性能，这些节点可以是虚拟机、裸机或专用设备，如 SG5712、SG5760、SG6060 或 SGF6024。在我们的测试中，我们使用 SG6060 设备以最小尺寸的三节点网格超出了 SmartStore 关键性能要求。当客户使用附加索引器扩展其 Splunk 基础设施时，他们可以添加更多存储节点来提高性能和容量。

负载均衡器和端点配置

StorageGRID中的管理节点提供网格管理器 UI（用户界面）和 REST API 端点来查看、配置和管理您的StorageGRID系统，以及审计日志来跟踪系统活动。为了为 Splunk SmartStore 远程存储提供高可用性 S3 端点，我们实施了StorageGRID负载均衡器，它作为管理节点和网关节点上的服务运行。此外，负载均衡器还管理本地流量并与 GSLB（全局服务器负载均衡）对话以帮助进行灾难恢复。

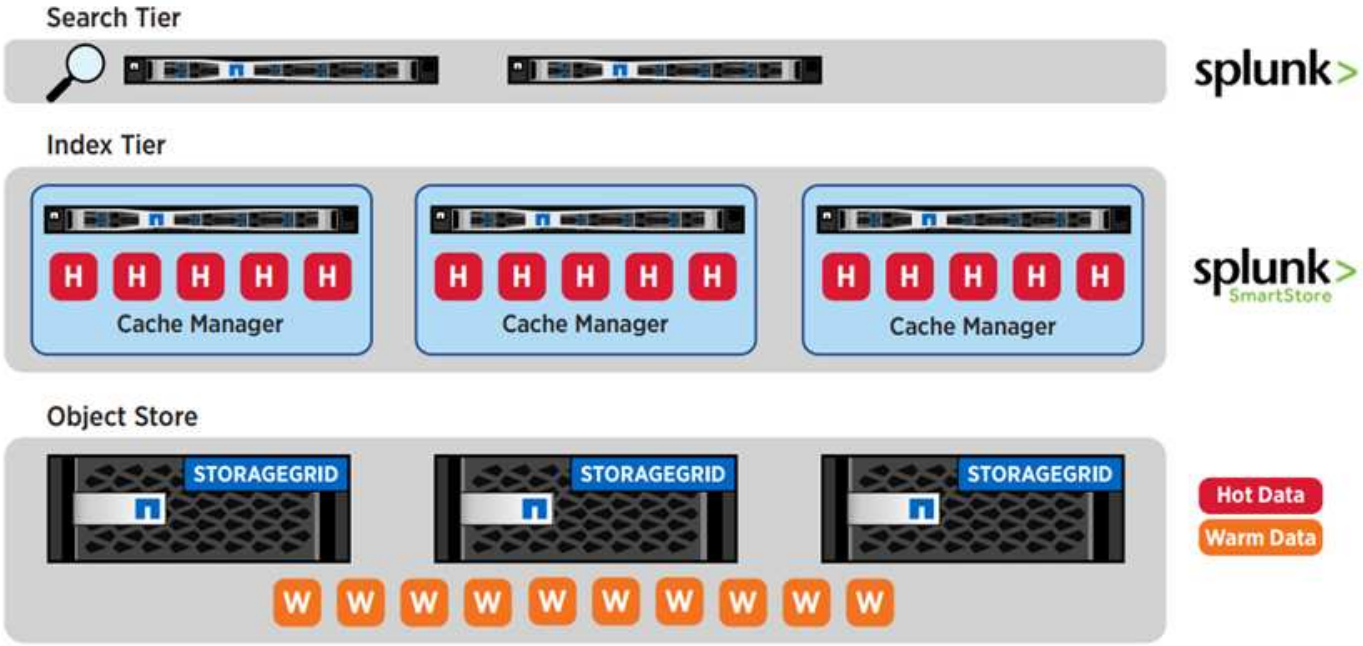
为了进一步增强端点配置， StorageGRID提供了内置于管理节点的流量分类策略，让您监控工作负载流量，并对工作负载应用各种服务质量 (QoS) 限制。流量分类策略应用于网关节点和管理节点的StorageGRID负载均衡器服务上的端点。这些策略可以帮助限制和监控流量。

智能分层和成本节约

当客户意识到 Splunk 数据分析的强大功能和易用性时，他们自然希望索引不断增长的数据量。随着数据量的增长，服务数据所需的计算和存储基础设施也在增长。由于旧数据的引用频率较低，因此投入相同数量的计算资源并消耗昂贵的主存储变得越来越低效。为了大规模运营，客户可以将热数据移动到更具成本效益的层，从而释放热数据的计算和主存储。

带有StorageGRID 的Splunk SmartStore 为组织提供了可扩展、高性能且经济高效的解决方案。由于 SmartStore 具有数据感知能力，它会自动评估数据访问模式，以确定哪些数据需要进行实时分析（热数据），哪些数据应该驻留在低成本的长期存储中（温数据）。 SmartStore 动态且智能地使用行业标准的 AWS S3 API ，将数据放置在StorageGRID提供的 S3 存储中。 StorageGRID灵活的横向扩展架构允许热数据层根据需要以经济高效的方式进行增长。 StorageGRID基于节点的架构确保性能和成本要求得到最佳满足。

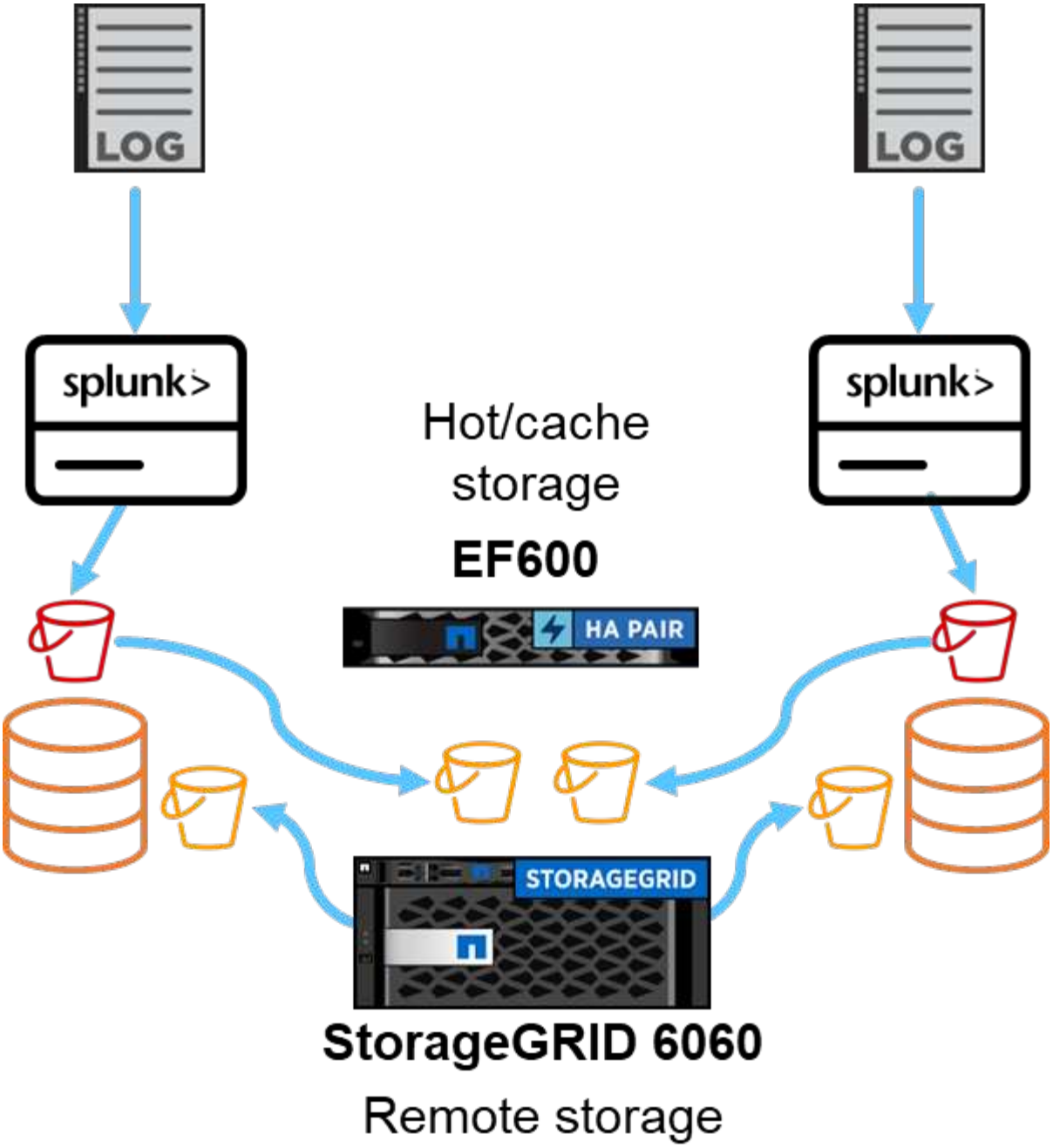
下图说明了 Splunk 和StorageGRID分层。



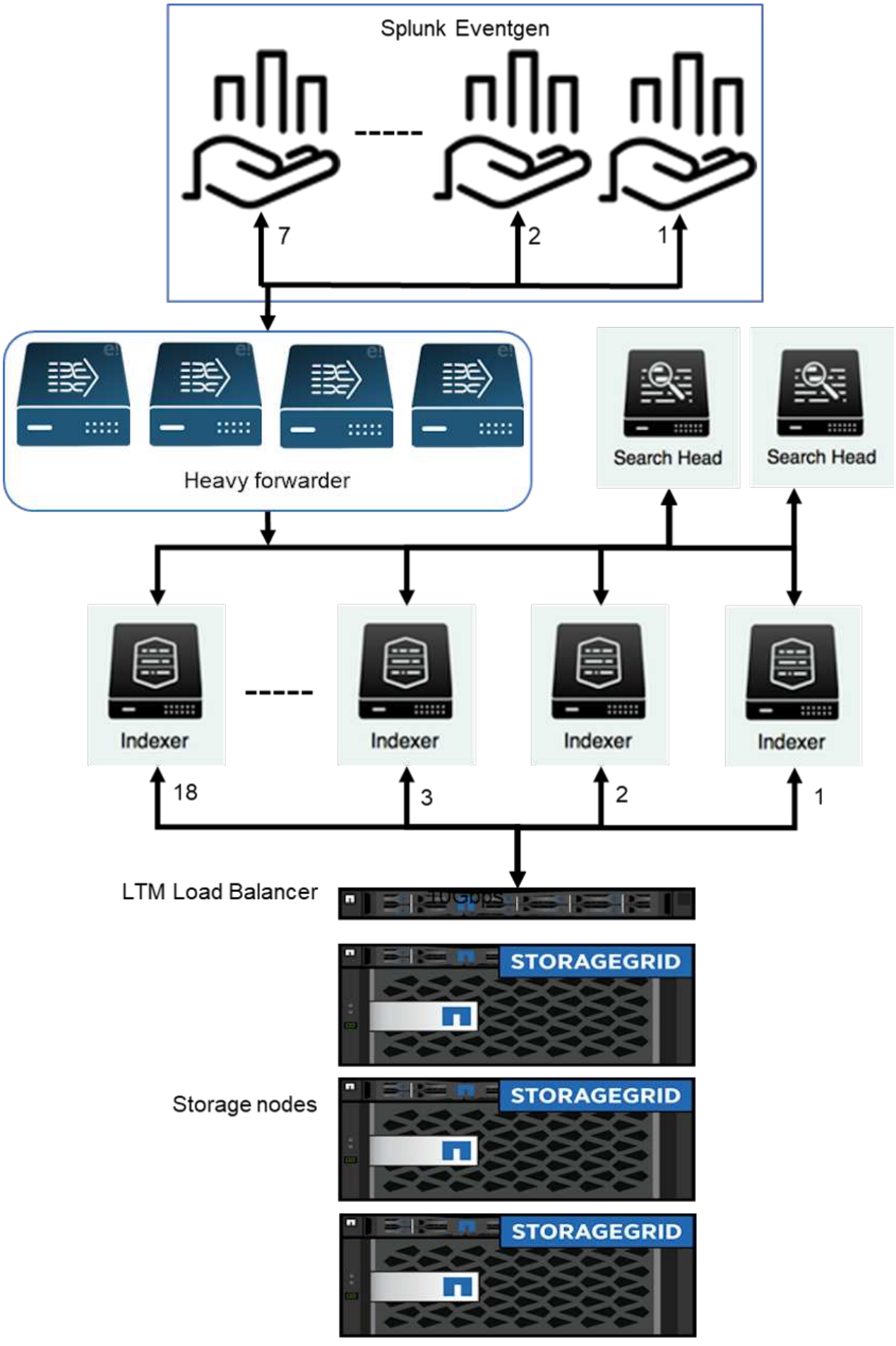
Splunk SmartStore 与NetApp StorageGRID的业界领先组合通过全栈解决方案提供了解耦架构的优势。

单站点 **SmartStore** 性能

本节介绍 Splunk SmartStore 在NetApp StorageGRID控制器上的性能。 Splunk SmartStore 将热数据移动到远程存储，在本例中是性能验证中的StorageGRID对象存储。



我们使用 EF600 作为热/缓存存储，使用StorageGRID 6060 作为远程存储。我们使用以下架构进行性能验证。我们使用了两个搜索头、四个重型转发器将数据转发到索引器、七个 Splunk 事件生成器（Eventgens）来生成实时数据，以及 18 个索引器来存储数据。


Splunk Eventgen
7
2
1
Heavy forwarder
Search Head
Search Head
Indexer
Indexer
Indexer
Indexer
18
3
2
1
LTM Load Balancer
10Gbps
STORAGEGRID
STORAGEGRID
STORAGEGRID
Storage nodes

配置

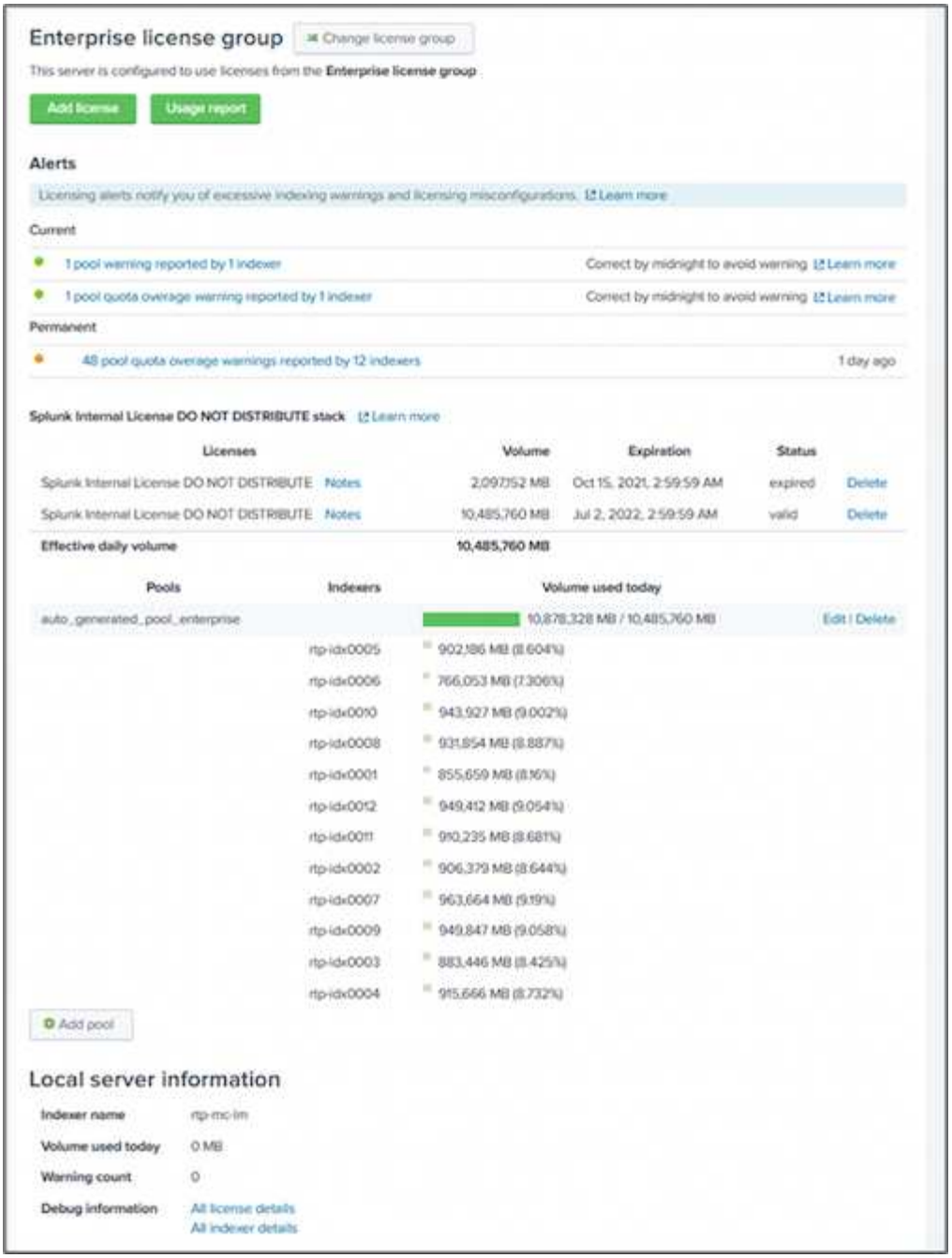
下表列出了用于 SmartStorage 性能验证的硬件。

Splunk 组件	任务	数量	核心	内存	操作系统
重型货运代理	负责提取数据并将数据转发给索引器	4	16 核	32 GB 内存	SLED 15 SP2
索引器	管理用户数据	18	16 核	32 GB 内存	SLED 15 SP2
搜索头	用户前端在索引器中搜索数据	2	16 核	32 GB 内存	SLED 15 SP2
搜索头部署器	处理搜索头集群的更新	1	16 核	32 GB 内存	SLED 15 SP2
集群主节点	管理 Splunk 安装和索引器	1	16 核	32 GB 内存	SLED 15 SP2
监控控制台和许可证主控器	对整个 Splunk 部署进行集中监控并管理 Splunk 许可证	1	16 核	32 GB 内存	SLED 15 SP2

SmartStore远程商店性能验证

在本次性能验证中，我们在所有索引器的本地存储中配置了 SmartStore 缓存，以保存 10 天的数据。我们启用了 `maxDataSize=auto`（750MB 存储桶大小）在 Splunk 集群管理器中并将更改推送到所有索引器。为了测量上传性能，我们在 10 天内每天摄取 10TB 的数据，并同时将所有热存储桶转为热存储桶，并从 SmartStore 监控控制台仪表板捕获每个实例和整个部署的峰值和平均吞吐量。

此图显示了一天内摄取的数据。



我们从集群主节点运行以下命令（索引名称是 `eventgen-test`）。然后，我们通过 SmartStore 监控控制台仪表板捕获每个实例和整个部署的峰值和平均上传吞吐量。

```
for i in rtp-idx0001 rtp-idx0002 rtp-idx0003 rtp-idx0004 rtp-idx0005 rtp-
idx0006 rtp-idx0007 rtp-idx0008 rtp-idx0009 rtp-idx0010 rtp-idx0011 rtp-
idx0012 rtp-idx0013011 rtdx0014 rtp-idx0015 rtp-idx0016 rtp-idx0017 rtp-
idx0018 ; do  ssh $i "hostname;  date; /opt/splunk/bin/splunk _internal
call /data/indexes/eventgen-test/roll-hot-buckets -auth admin:12345678;
sleep 1  "; done
```

集群主控对所有索引器（rtp-idx0001…rtp-idx0018）均采用无密码身份验证。

为了测量下载性能，我们使用以下命令运行两次 evict CLI，从缓存中逐出所有数据。

> 我们从集群主机运行以下命令，并从搜索头基于来自StorageGRID的远程存储的 10 天数据运行搜索。然后，我们通过 SmartStore 监控控制台仪表板捕获每个实例和整个部署的峰值和平均上传吞吐量。

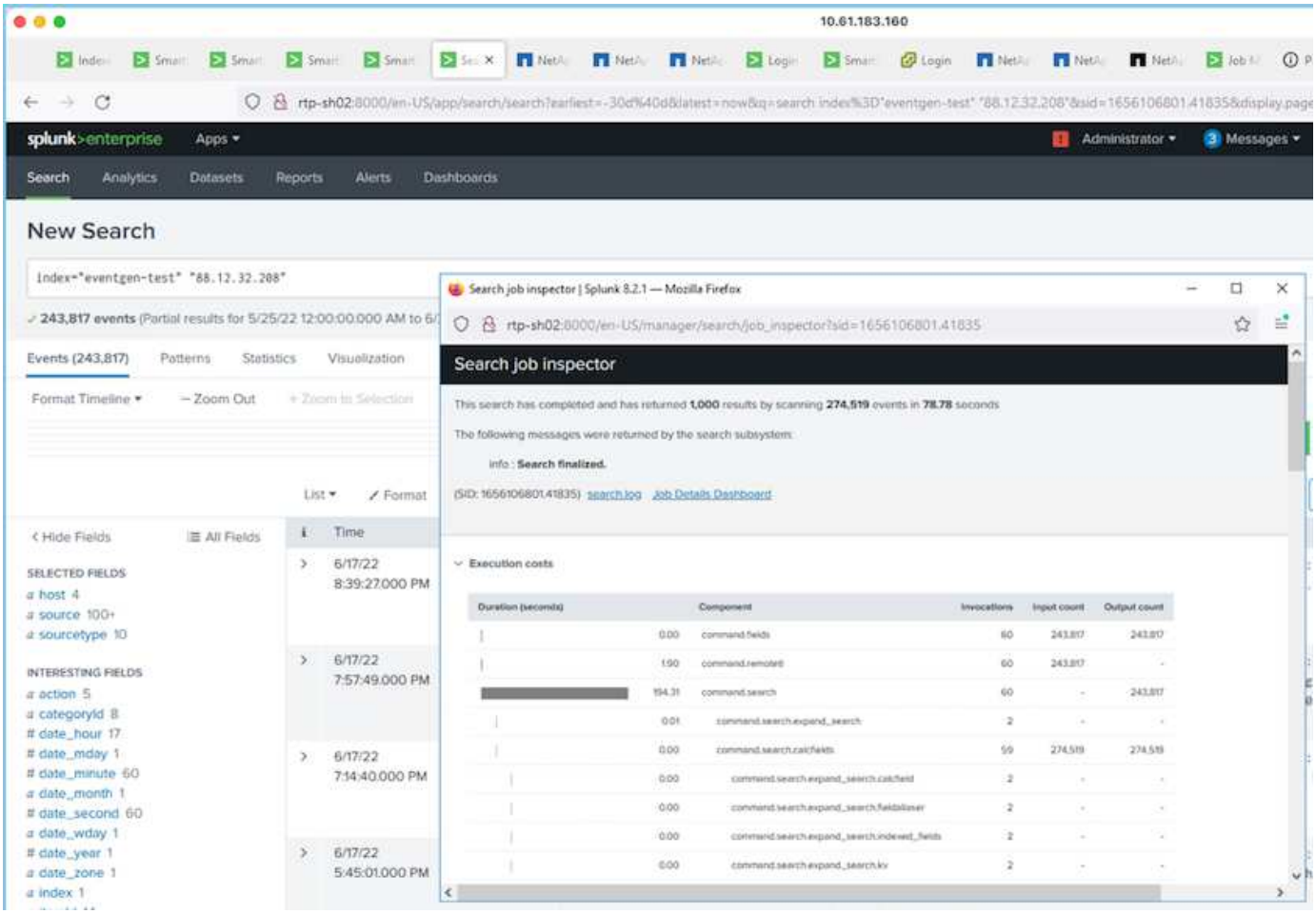
```
for i in rtp-idx0001 rtp-idx0002 rtp-idx0003 rtp-idx0004 rtp-idx0005 rtp-
idx0006 rtp-idx0007 rtp-idx0008 rtp-idx0009 rtp-idx0010 rtp-idx0011 rtp-
idx0012 rtp-idx0013 rtp-idx0014 rtp-idx0015 rtp-idx0016 rtp-idx0017 rtp-
idx0018 ; do  ssh $i " hostname;  date; /opt/splunk/bin/splunk _internal
call /services/admin/cacheman/_evict -post:mb 1000000000 -post:path
/mnt/EF600 -method POST  -auth admin:12345678;   "; done
```

索引器配置是从 SmartStore 集群主机推送的。集群主机对索引器有以下配置。

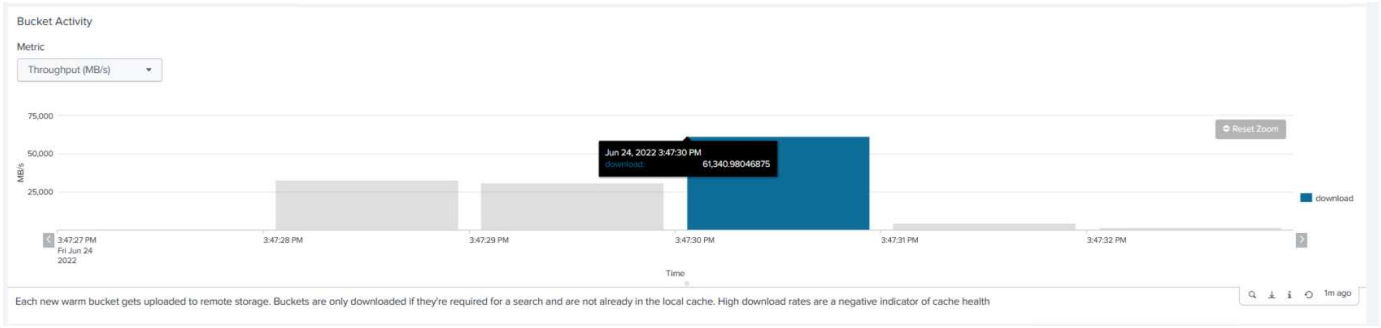
```
Rtp-cm01:~ # cat /opt/splunk/etc/master-apps/_cluster/local/indexes.conf
[default]
maxDataSize = auto
#defaultDatabase = eventgen-basic
defaultDatabase = eventgen-test
hotlist_recency_secs = 864000
repFactor = auto
[volume:remote_store]
storageType = remote
path = s3://smartstore2
remote.s3.access_key = U64TUHONBNC98GQGL60R
remote.s3.secret_key = UBoXNE0jmECie05Z7iCYVzbSB6WJFckiYLcdm2yg
remote.s3.endpoint = 3.sddc.netapp.com:10443
remote.s3.signature_version = v2
remote.s3.clientCert =
[eventgen-basic]
homePath = $SPLUNK_DB/eventgen-basic/db
coldPath = $SPLUNK_DB/eventgen-basic/colddb
thawedPath = $SPLUNK_DB/eventgen-basic/thawed
[eventgen-migration]
homePath = $SPLUNK_DB/eventgen-scale/db
coldPath = $SPLUNK_DB/eventgen-scale/colddb
thawedPath = $SPLUNK_DB/eventgen-scale/thaweddb
[main]
homePath = $SPLUNK_DB/$_index_name/db
coldPath = $SPLUNK_DB/$_index_name/colddb
thawedPath = $SPLUNK_DB/$_index_name/thaweddb
[history]
homePath = $SPLUNK_DB/$_index_name/db
coldPath = $SPLUNK_DB/$_index_name/colddb
```

```
thawedPath = $SPLUNK_DB/$_index_name/thaweddb
[summary]
homePath = $SPLUNK_DB/$_index_name/db
coldPath = $SPLUNK_DB/$_index_name/colddb
thawedPath = $SPLUNK_DB/$_index_name/thaweddb
[remote-test]
homePath = $SPLUNK_DB/$_index_name/db
coldPath = $SPLUNK_DB/$_index_name/colddb
#for storagegrid config
remotePath = volume:remote_store/$_index_name
thawedPath = $SPLUNK_DB/$_index_name/thaweddb
[eventgen-test]
homePath = $SPLUNK_DB/$_index_name/db
maxDataSize=auto
maxHotBuckets=1
maxWarmDBCount=2
coldPath = $SPLUNK_DB/$_index_name/colddb
#for storagegrid config
remotePath = volume:remote_store/$_index_name
thawedPath = $SPLUNK_DB/$_index_name/thaweddb
[eventgen-evict-test]
homePath = $SPLUNK_DB/$_index_name/db
coldPath = $SPLUNK_DB/$_index_name/colddb
#for storagegrid config
remotePath = volume:remote_store/$_index_name
thawedPath = $SPLUNK_DB/$_index_name/thaweddb
maxDataSize = auto_high_volume
maxWarmDBCount = 5000
rtp-cm01:~ #
```

我们在搜索头上运行了以下搜索查询来收集性能矩阵。



我们从集群主机收集了性能信息。峰值性能为61.34GBps。



平均性能约为 29GBps。

StorageGRID性能

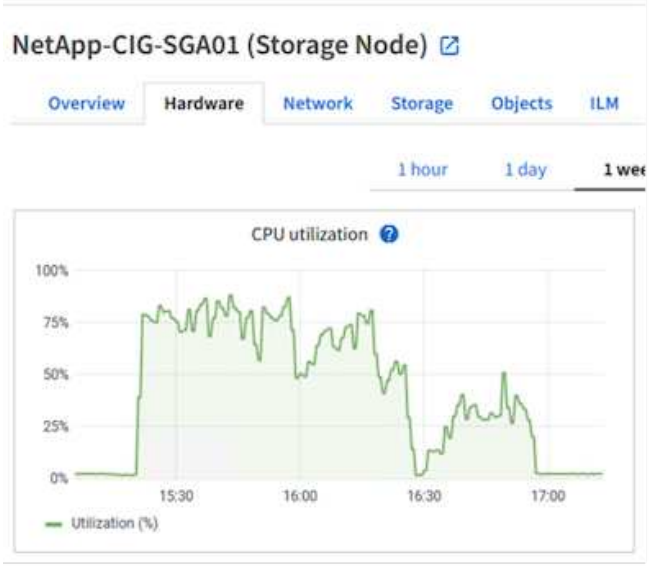
SmartStore 的性能基于从大量数据中搜索特定的模式和字符串。在此验证中，事件是使用 "事件生成"通过搜索头在特定的 Splunk 索引（eventgen-test）上进行搜索，并且请求对于大多数查询转到StorageGRID 。下图显示了查询数据的命中和未命中情况。命中数据来自本地磁盘，未命中数据来自StorageGRID控制器。

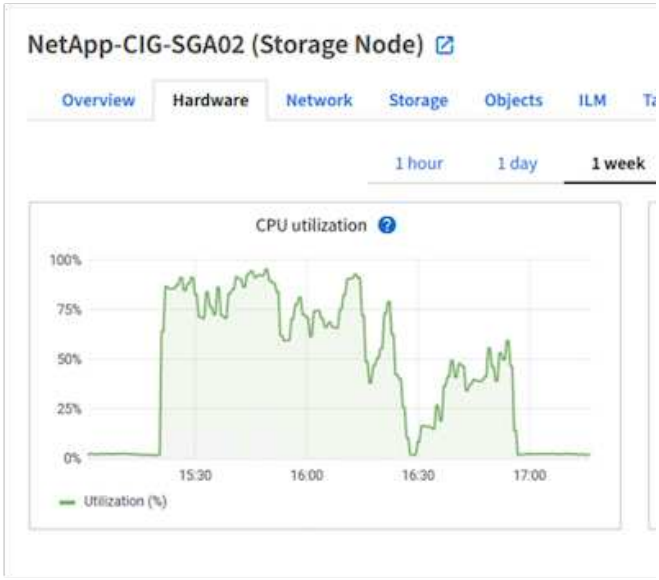
绿色显示命中数据，橙色显示未命中数据。

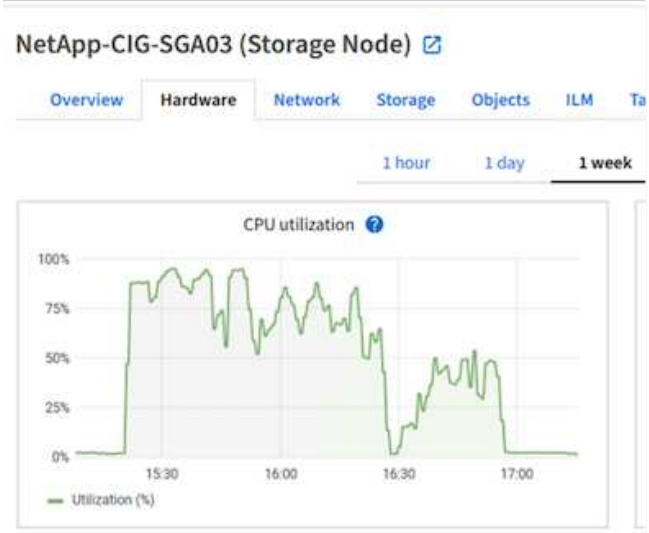
当在StorageGRID上运行搜索查询时， StorageGRID的 S3 检索率的时间如下图所示。

StorageGRID硬件使用情况

StorageGRID实例有一个负载均衡器和三个StorageGRID控制器。所有三个控制器的 CPU 利用率均为 75% 至 100%。







采用**NetApp**存储控制器的 **SmartStore -** 为客户带来好处

- 将计算和存储分离。 Splunk SmartStore 将计算和存储分离，帮助您独立扩展它们。
- 按需提供数据。 SmartStore 使数据接近按需计算，并提供计算和存储弹性和成本效率，以实现更长时间的大规模数据保留。
- 符合 **AWS S3 API**。 SmartStore 使用 AWS S3 API 与恢复存储进行通信，恢复存储是符合 AWS S3 和 S3 API 的对象存储，例如StorageGRID。
- 减少存储需求和成本。 SmartStore 减少了老化数据（暖/冷）的存储要求。它只需要一份数据副本，因为NetApp存储提供数据保护并处理故障和高可用性。
- 硬件故障。 SmartStore 部署中的节点故障不会导致数据无法访问，并且索引器从硬件故障或数据不平衡中恢复的速度更快。
- 应用程序和数据感知缓存。
- 按需添加或删除索引器以及设置或拆除集群。
- 存储层不再与硬件相关。

结束语

Splunk Enterprise 是市场领先的 SIEM 解决方案，可推动安全、IT 和 DevOps 团队取得成果。我们客户组织中 Splunk 的使用量已显著增加。因此，需要添加更多数据源，同时保留更长时间的数据，从而给 Splunk 基础设施带来压力。

Splunk SmartStore 和NetApp StorageGRID的结合旨在为组织提供可扩展的架构，以通过 SmartStore 和StorageGRID对象存储实现更高的摄取性能，并提高跨多个地理区域的 Splunk 环境的可扩展性。

在哪里可以找到更多信息

要了解有关本文档中描述的信息的更多信息，请查看以下文档和/或网站：

- "NetApp StorageGRID文档资源"
- "NetApp产品文档"
- "Splunk Enterprise 文档"
- "Splunk Enterprise 关于 SmartStore"
- "Splunk Enterprise 分布式部署手册"
- "Splunk Enterprise 管理索引器和索引器集群"

版权信息

版权所有 © 2025 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。