NetApp StorageGRID con Splunk SmartStore

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Tabla de contenidos

NetApp StorageGRID con Splunk SmartStore

TR-4869: NetApp StorageGRID con Splunk SmartStore

Splunk Enterprise es la solución de gestión de eventos e información de seguridad (SIEM) líder en el mercado que impulsa resultados en los equipos de seguridad, TI y DevOps.

Descripción general

Los volúmenes de datos continúan creciendo a un ritmo exponencial, lo que crea enormes oportunidades para las empresas que pueden aprovechar este vasto recurso. Splunk Enterprise continúa ganando adopción en una variedad más amplia de casos de uso. A medida que crecen los casos de uso, también crece la cantidad de datos que Splunk Enterprise ingiere y procesa. La arquitectura tradicional de Splunk Enterprise es un diseño de escalamiento distribuido que proporciona excelente acceso y disponibilidad de datos. Sin embargo, las empresas que utilizan esta arquitectura se enfrentan a costos crecientes asociados con la escalabilidad para satisfacer el rápido crecimiento del volumen de datos.

Splunk SmartStore con NetApp StorageGRID resuelve este desafío al ofrecer un nuevo modelo de implementación en el que el cómputo y el almacenamiento están desacoplados. Esta solución también desbloquea una escala y elasticidad inigualables para los entornos de Splunk Enterprise al permitir a los clientes escalar entre sitios únicos y múltiples, al mismo tiempo que reduce los costos al permitir que el cómputo y el almacenamiento escalen de forma independiente y agrega niveles inteligentes al almacenamiento de objetos S3 basado en la nube rentable.

La solución optimiza la cantidad de datos en el almacenamiento local mientras mantiene el rendimiento de la búsqueda, lo que permite escalar el cómputo y el almacenamiento según demanda. SmartStore evalúa automáticamente los patrones de acceso a los datos para determinar qué datos deben ser accesibles para realizar análisis en tiempo real y qué datos deben residir en el almacenamiento de objetos S3 de menor costo.

Este informe técnico describe el beneficio que NetApp brinda a una solución Splunk SmartStore y al mismo tiempo demuestra un marco para diseñar y dimensionar Splunk SmartStore en su entorno. El resultado es una solución sencilla, escalable y resistente que ofrece un coste total de propiedad atractivo. StorageGRID proporciona almacenamiento de objetos escalable y rentable basado en el protocolo S3/API, también conocido como almacenamiento remoto, lo que permite a las organizaciones escalar su solución Splunk a un menor costo y al mismo tiempo aumentar la resiliencia.

Splunk SmartStore se refiere al almacenamiento de objetos como almacenes remotos o niveles de almacenamiento remoto.

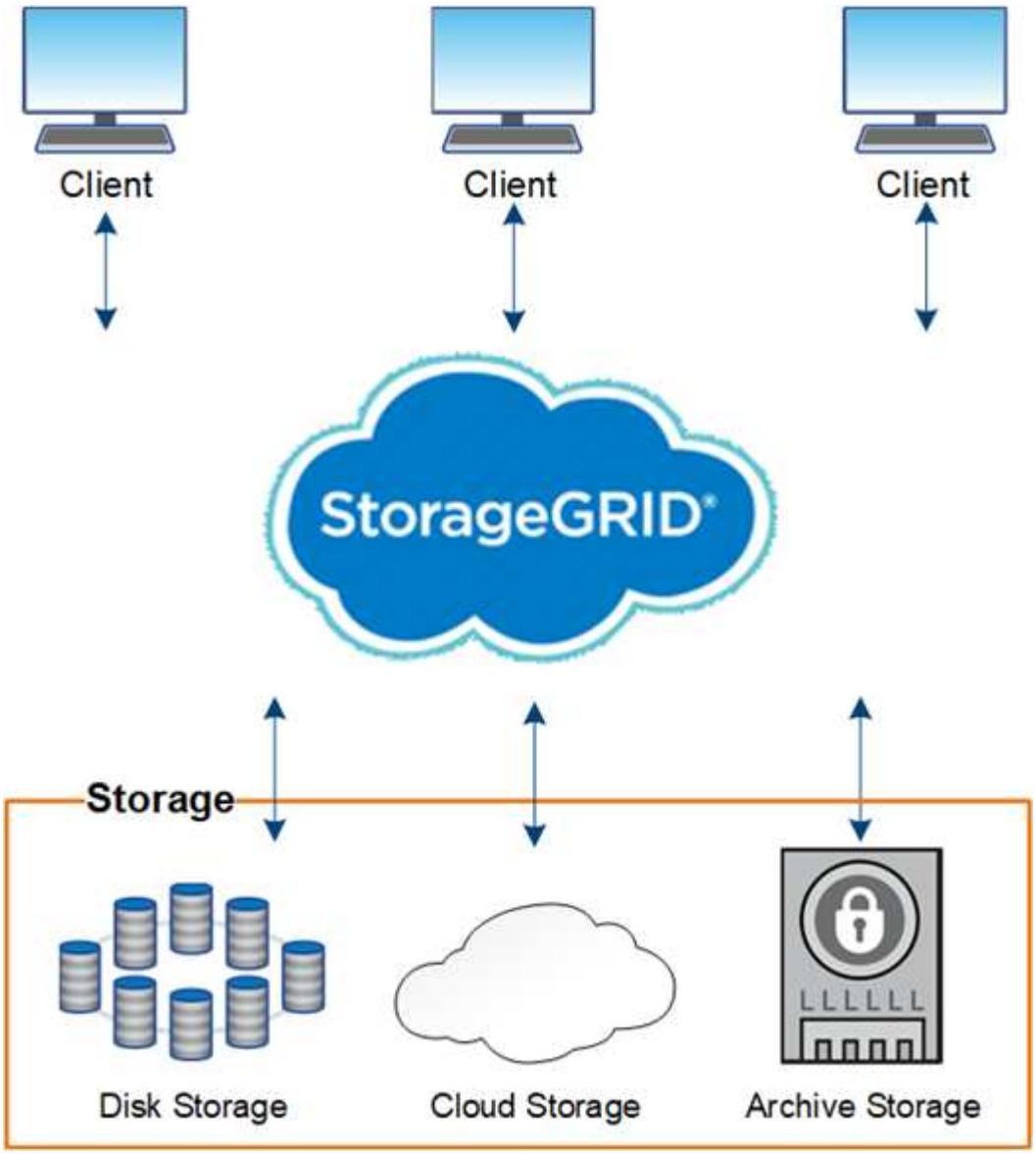
Acerca de NetApp StorageGRID

NetApp StorageGRID es una solución de almacenamiento de objetos definida por software para archivos grandes, repositorios de medios y almacenes de datos web. Con StorageGRID, NetApp aprovecha dos décadas de experiencia en la entrega de soluciones de innovación y gestión de datos líderes en la industria, al tiempo que ayuda a las organizaciones a administrar y maximizar el valor de su información tanto en sus instalaciones como en implementaciones de nube pública, privada o híbrida.

StorageGRID proporciona almacenamiento seguro y duradero para datos no estructurados a escala. Las políticas de gestión del ciclo de vida integradas y basadas en metadatos optimizan dónde residen sus datos a lo largo de su vida. El contenido se coloca en el lugar correcto, en el momento correcto y en el nivel de almacenamiento correcto para reducir costos. El espacio de nombres único permite acceder a los datos a

través de una única llamada, independientemente de la ubicación geográfica del almacenamiento StorageGRID . Los clientes pueden implementar y administrar múltiples instancias de StorageGRID entre centros de datos y en la infraestructura de la nube.

Un sistema StorageGRID está compuesto de nodos heterogéneos, redundantes y distribuidos globalmente que pueden integrarse con aplicaciones cliente existentes y de próxima generación.



IDC MarketScape nombró recientemente a NetApp como líder en el último informe, IDC MarketScape: Evaluación de proveedores de almacenamiento basado en objetos a nivel mundial 2019. Con casi 20 años de implementaciones de producción en las industrias más exigentes, StorageGRID es un líder reconocido en datos no estructurados.

Con StorageGRID, puede lograr lo siguiente:

- Implemente múltiples instancias de StorageGRID para acceder a datos desde cualquier ubicación entre centros de datos y la nube a través de un único espacio de nombres que escala fácilmente a cientos de petabytes.
- Proporciona flexibilidad para implementar y administrar de forma centralizada todas las infraestructuras.
- Proporciona una durabilidad inigualable con quince nueves de durabilidad aprovechando la codificación de

borrado en capas (EC).

- Habilite más capacidades de múltiples nubes híbridas con integraciones validadas en Amazon S3 Glacier y Azure Blob.
- Cumpla con las obligaciones regulatorias y facilite el cumplimiento mediante la retención de datos a prueba de manipulaciones, sin API propietarias ni dependencia de proveedores.

Para obtener más información sobre cómo StorageGRID puede ayudarlo a resolver sus problemas de administración de datos no estructurados más complejos, consulte "Página de inicio de NetApp StorageGRID" .

Acerca de Splunk Enterprise

Splunk Enterprise es una plataforma para convertir datos en acciones. Los datos generados por varias fuentes, como archivos de registro, sitios web, dispositivos, sensores y aplicaciones, se envían y analizan en los indexadores de Splunk, lo que le permite obtener información valiosa de los datos. Puede identificar violaciones de datos, señalar tendencias de clientes y productos, encontrar oportunidades para optimizar la infraestructura o crear información útil en una amplia variedad de casos de uso.

Acerca de Splunk SmartStore

Splunk SmartStore amplía los beneficios de la arquitectura de Splunk al tiempo que simplifica su capacidad de escalar de manera rentable. La disociación de los recursos de cómputo y almacenamiento da como resultado nodos de indexación optimizados para E/S con necesidades de almacenamiento significativamente reducidas porque solo almacenan un subconjunto de datos como caché. No es necesario agregar procesamiento o almacenamiento adicional cuando solo es necesario uno de esos recursos, lo que le permite obtener ahorros de costos significativos. Puede utilizar un almacenamiento de objetos basado en S3 rentable y fácilmente escalable, que simplifica aún más el entorno, reduce los costos y le permite mantener un conjunto de datos más masivo.

Splunk SmartStore ofrece un valor significativo a las organizaciones, incluido lo siguiente:

- Reducción de los costos de almacenamiento al trasladar datos no utilizados a un almacenamiento de objetos S3 con costos optimizados
- Escalabilidad fluida mediante la separación del almacenamiento y la computación
- Simplificar la continuidad del negocio aprovechando el almacenamiento nativo de la nube resiliente

Descripción general de la solución

Esta página describe los componentes utilizados para completar esta solución, incluidos NetApp StorageGRID, Splunk Enterprise y Splunk SmartStore.

StorageGRID en NetApp

NetApp StorageGRID es una plataforma de almacenamiento de objetos de alto rendimiento y rentable. Ofrece una gestión de datos global inteligente e impulsada por políticas utilizando una arquitectura de red distribuida basada en nodos. Simplifica la gestión de petabytes de datos no estructurados y miles de millones de objetos a través de su omnipresente espacio de nombres de objetos globales combinado con sofisticadas funciones de gestión de datos. El acceso a objetos mediante una sola llamada se extiende a través de los sitios y simplifica las arquitecturas de alta disponibilidad al tiempo que garantiza el acceso continuo a los objetos independientemente de las interrupciones del sitio o la infraestructura.

La multitenencia permite que múltiples aplicaciones de datos no estructurados empresariales y en la nube se administren de forma segura dentro de la misma red, lo que aumenta el ROI y los casos de uso de StorageGRID. Se pueden crear múltiples niveles de servicio con políticas de ciclo de vida de objetos basadas en metadatos, optimizando la durabilidad, la protección, el rendimiento y la localidad en múltiples geografías. Los usuarios pueden ajustar las políticas y realinear el panorama de datos sin interrupciones a medida que cambian sus requisitos.

SmartStore aprovecha StorageGRID como nivel de almacenamiento remoto y permite a los clientes implementar múltiples sitios distribuidos geográficamente para lograr una disponibilidad y durabilidad sólidas, presentadas como un único espacio de nombres de objetos. Esto permite que Splunk SmartStore aproveche el alto rendimiento, la capacidad densa y la capacidad de escalar a cientos de nodos en múltiples sitios físicos utilizando una única URL para interactuar con los objetos. Esta URL única también permite que la expansión del almacenamiento, las actualizaciones y las reparaciones no produzcan interrupciones, incluso más allá de un solo sitio. El motor de políticas de gestión de datos exclusivo de StorageGRID proporciona niveles optimizados de rendimiento y durabilidad y cumplimiento de los requisitos de localidad de datos.

Splunk Enterprise

Splunk, líder en la recopilación y análisis de datos generados por máquinas, ayuda a simplificar y modernizar la TI a través de sus capacidades de análisis operativo. También se expande a casos de uso de análisis de negocios, seguridad e IoT. El almacenamiento es un elemento fundamental para una implementación exitosa del software Splunk.

Los datos generados por máquinas son el tipo de big data de más rápido crecimiento. El formato es impredecible y proviene de muchas fuentes diferentes, a menudo a gran velocidad y en grandes volúmenes. Estas características de carga de trabajo a menudo se denominan escape digital. Splunk SmartStore ayuda a dar sentido a estos datos y proporciona una clasificación inteligente de datos para una ubicación optimizada de datos calientes y tibios en el nivel de almacenamiento más rentable.

Tienda inteligente de Splunk

Splunk SmartStore es una capacidad de indexación que utiliza almacenamiento de objetos (también conocido como almacenamiento remoto o niveles de almacenamiento remoto) como StorageGRID para almacenar datos cálidos mediante el protocolo S3.

A medida que aumenta el volumen de datos de una implementación, la demanda de almacenamiento generalmente supera la demanda de recursos informáticos. SmartStore le permite administrar su almacenamiento de indexador y sus recursos computacionales de manera rentable al escalar el procesamiento y el almacenamiento por separado.

SmartStore presenta un nivel de almacenamiento remoto, utilizando el protocolo S3, y un administrador de caché. Estas características permiten que los datos residan localmente en indexadores o en almacenamiento remoto. El administrador de caché, que reside en el indexador, administra el movimiento de datos entre el indexador y el nivel de almacenamiento remoto. Los datos se almacenan en depósitos (calientes y templados) junto con los metadatos del depósito.

Con SmartStore, puede reducir el espacio de almacenamiento del indexador al mínimo y elegir recursos informáticos optimizados para E/S porque la mayoría de los datos residen en el nivel de almacenamiento remoto. El indexador mantiene un caché local, que representa la cantidad mínima de datos necesarios para devolver los resultados solicitados y predichos. La caché local contiene buckets activos, copias de buckets cálidos que participan en búsquedas activas o recientes y metadatos de buckets.

Splunk SmartStore con StorageGRID permite a los clientes escalar de forma incremental el entorno con almacenamiento remoto rentable y de alto rendimiento al tiempo que proporciona un alto grado de elasticidad a la solución general. Esto permite a los clientes agregar cualquier componente (almacenamiento activo y/o

almacenamiento S3 templado) en cualquier cantidad determinada en cualquier momento, ya sea que necesiten más indexadores, cambiar la retención de datos o aumentar la tasa de ingesta sin ninguna interrupción.

Beneficios de esta solución

La solución permite agregar recursos de cómputo, almacenamiento activo o S3 para satisfacer la creciente demanda en términos de cantidad de usuarios o tasa de ingesta en implementaciones de sitios únicos o múltiples.

- **Actuación.** La combinación de Splunk SmartStore y NetApp StorageGRID proporciona una migración rápida de datos entre contenedores activos y contenedores tibios mediante almacenamiento de objetos. StorageGRID potencia el proceso de migración al proporcionar un rendimiento rápido para cargas de trabajo de objetos grandes.
- **Listo para múltiples sitios.** La arquitectura distribuida de StorageGRID permite a Splunk SmartStore extender las implementaciones en sitios individuales y múltiples a través de un único espacio de nombres global donde se puede acceder a los datos desde cualquier sitio, independientemente de dónde se encuentren.
- **Escalabilidad mejorada.** Escale los recursos de almacenamiento independientemente de los recursos computacionales para satisfacer las necesidades y demandas cambiantes en su entorno Splunk, proporcionando así un mejor TCO.
- **Capacidad.** Afronte el rápido crecimiento de los volúmenes en la implementación de Splunk con StorageGRID escalando un único espacio de nombres a más de 560 PB.
- **Disponibilidad de datos.** Optimice la disponibilidad de datos, el rendimiento, la distribución geográfica, la retención, la protección y los costos de almacenamiento con políticas basadas en metadatos que pueden ajustarse dinámicamente a medida que evoluciona el valor comercial de sus datos.

Aumente el rendimiento con el caché SmartStore, que es un componente del indexador que maneja la transferencia de copias de bucket entre el almacenamiento local (activo) y el remoto (tibio). El dimensionamiento de Splunk para esta solución se basa en "Pautas proporcionadas por Splunk" . La solución permite agregar recursos de cómputo, almacenamiento activo o S3 para satisfacer la creciente demanda en términos de cantidad de usuarios o tasa de ingesta en implementaciones de sitios únicos o múltiples.

Arquitectura de Splunk

Esta sección describe la arquitectura de Splunk, incluidas las definiciones clave, las implementaciones distribuidas de Splunk, Splunk SmartStore, el flujo de datos, los requisitos de hardware y software, los requisitos de sitios únicos y múltiples, etc.

Definiciones clave

Las siguientes dos tablas enumeran los componentes de Splunk y NetApp utilizados en la implementación distribuida de Splunk.

Esta tabla enumera los componentes de hardware de Splunk para la configuración distribuida de Splunk Enterprise.

Componente de Splunk	Tarea
Indexador	Repositorio de datos de Splunk Enterprise

Componente de Splunk	Tarea
Reenvío universal	Responsable de ingerir datos y enviarlos a los indexadores.
Cabezal de búsqueda	La interfaz de usuario utilizada para buscar datos en los indexadores
Maestro del clúster	Administra la instalación de indexadores y cabezales de búsqueda de Splunk
Consola de monitoreo	Herramienta de monitoreo centralizada utilizada en toda la implementación
Maestro de licencias	El administrador de licencias gestiona las licencias de Splunk Enterprise
Servidor de implementación	Actualiza las configuraciones y distribuye aplicaciones al componente de procesamiento
Componente de almacenamiento	Tarea
AFF de NetApp	Almacenamiento totalmente flash utilizado para administrar datos de nivel activo. También conocido como almacenamiento local.
StorageGRID en NetApp	Almacenamiento de objetos S3 utilizado para administrar datos de nivel cálido. SmartStore lo utiliza para mover datos entre el nivel caliente y el nivel templado. También conocido como almacenamiento remoto.

Esta tabla enumera los componentes de la arquitectura de almacenamiento de Splunk.

Componente de Splunk	Tarea	Componente responsable
Tienda inteligente	Proporciona a los indexadores la capacidad de organizar datos en niveles desde el almacenamiento local hasta el almacenamiento de objetos.	Splunk
Caliente	El lugar de aterrizaje donde los reenvíos universales colocan los datos recién escritos. El almacenamiento se puede escribir y los datos se pueden buscar. Este nivel de datos normalmente está compuesto por SSD o discos duros rápidos.	ONTAP
Administrador de caché	Administra el caché local de datos indexados, recupera datos confidenciales del almacenamiento remoto cuando se realiza una búsqueda y expulsa del caché los datos menos utilizados.	Tienda inteligente

Componente de Splunk	Tarea	Componente responsable
Cálido	Los datos se transfieren de manera lógica al depósito y se renombran primero al nivel cálido desde el nivel caliente. Los datos dentro de este nivel están protegidos y, al igual que el nivel activo, pueden estar compuestos por SSD o HDD de mayor capacidad. Se admiten copias de seguridad incrementales y completas mediante soluciones de protección de datos comunes.	StorageGRID

Implementaciones distribuidas de Splunk

Para dar soporte a entornos más grandes en los que los datos se originan en muchas máquinas, es necesario procesar grandes volúmenes de datos. Si muchos usuarios necesitan buscar datos, puede escalar la implementación distribuyendo instancias de Splunk Enterprise en varias máquinas. Esto se conoce como una implementación distribuida.

En una implementación distribuida típica, cada instancia de Splunk Enterprise realiza una tarea especializada y reside en uno de los tres niveles de procesamiento correspondientes a las funciones de procesamiento principales.

La siguiente tabla enumera los niveles de procesamiento de Splunk Enterprise.

Nivel	Componente	Descripción
Entrada de datos	Promotor	Un reenvío consume datos y luego los reenvía a un grupo de indexadores.
Indexación	Indexador	Un indexador indexa los datos entrantes que normalmente recibe de un grupo de reenvíos. El indexador transforma los datos en eventos y almacena los eventos en un índice. El indexador también busca los datos indexados en respuesta a las solicitudes de búsqueda de un cabezal de búsqueda.
Gestión de búsquedas	Cabezal de búsqueda	Un cabezal de búsqueda sirve como recurso central para la búsqueda. Los cabezales de búsqueda de un clúster son intercambiables y tienen acceso a las mismas búsquedas, paneles, objetos de conocimiento, etc., desde cualquier miembro del clúster de cabezales de búsqueda.

En la siguiente tabla se enumeran los componentes importantes que se utilizan en un entorno distribuido de

Splunk Enterprise.

Componente	Descripción	Responsabilidad
Maestro del clúster de índices	Coordina actividades y actualizaciones de un clúster de indexadores	Gestión de índices
Clúster de índices	Grupo de indexadores de Splunk Enterprise que están configurados para replicar datos entre sí	Indexación
Desplegador de cabezal de búsqueda	Maneja la implementación y las actualizaciones del maestro del clúster.	Gestión de cabezales de búsqueda
Grupo de búsqueda de cabezas	Grupo de cabezales de búsqueda que sirve como recurso central para la búsqueda	Gestión de búsquedas
Balanceadores de carga	Lo utilizan los componentes agrupados para gestionar la creciente demanda de los cabezales de búsqueda, los indexadores y el destino S3 para distribuir la carga entre los componentes agrupados.	Gestión de carga para componentes agrupados

Vea los siguientes beneficios de las implementaciones distribuidas de Splunk Enterprise:

- Acceder a fuentes de datos diversas o dispersas
- Proporcionar funcionalidad para gestionar las necesidades de datos de empresas de cualquier tamaño y complejidad.
- Logre alta disponibilidad y garantice la recuperación ante desastres con replicación de datos e implementación en múltiples sitios

Tienda inteligente de Splunk

SmartStore es una capacidad de indexación que permite que los almacenes de objetos remotos, como Amazon S3, almacenen datos indexados. A medida que aumenta el volumen de datos de una implementación, la demanda de almacenamiento generalmente supera la demanda de recursos computacionales. SmartStore le permite administrar su almacenamiento de indexador y sus recursos computacionales de manera rentable al escalar esos recursos por separado.

SmartStore presenta un nivel de almacenamiento remoto y un administrador de caché. Estas características permiten que los datos residan localmente en indexadores o en el nivel de almacenamiento remoto. El administrador de caché administra el movimiento de datos entre el indexador y el nivel de almacenamiento remoto, que está configurado en el indexador.

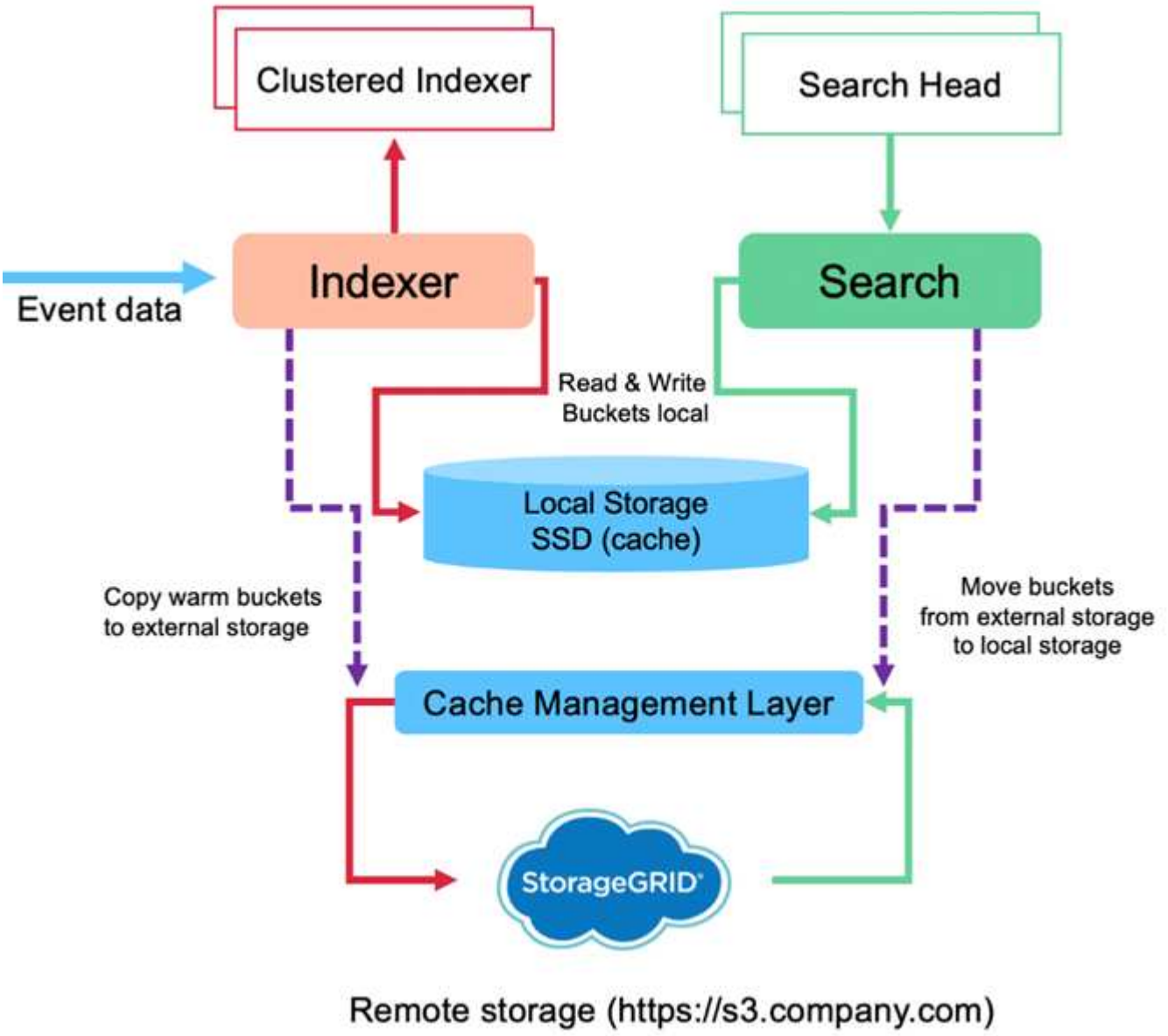
Con SmartStore, puede reducir el espacio de almacenamiento del indexador al mínimo y elegir recursos informáticos optimizados para E/S. La mayoría de los datos residen en el almacenamiento remoto. El indexador mantiene un caché local que contiene una cantidad mínima de datos: buckets activos, copias de buckets cálidos que participan en búsquedas activas o recientes y metadatos de buckets.

Flujo de datos de Splunk SmartStore

Cuando los datos provenientes de varias fuentes llegan a los indexadores, estos se indexan y se guardan localmente en un contenedor activo. El indexador también replica los datos del contenedor activo en los indexadores de destino. Hasta ahora, el flujo de datos es idéntico al flujo de datos de los índices que no son SmartStore.

Cuando el cubo caliente se calienta demasiado, el flujo de datos diverge. El indexador de origen copia el depósito cálido en el almacén de objetos remoto (nivel de almacenamiento remoto) mientras deja la copia existente en su caché, porque las búsquedas tienden a ejecutarse en datos indexados recientemente. Sin embargo, los indexadores de destino eliminan sus copias porque el almacén remoto proporciona alta disponibilidad sin mantener múltiples copias locales. La copia maestra del depósito ahora reside en el almacén remoto.

La siguiente imagen muestra el flujo de datos de Splunk SmartStore.



El administrador de caché del indexador es fundamental para el flujo de datos de SmartStore. Obtiene copias

de depósitos del almacén remoto según sea necesario para manejar solicitudes de búsqueda. También expulsa del caché copias de los buckets más antiguas o menos buscadas, porque la probabilidad de que participen en búsquedas disminuye con el tiempo.

El trabajo del administrador de caché es optimizar el uso del caché disponible y garantizar que las búsquedas tengan acceso inmediato a los segmentos que necesitan.

Requisitos de software

La siguiente tabla enumera los componentes de software necesarios para implementar la solución. Los componentes de software que se utilizan en cualquier implementación de la solución pueden variar según los requisitos del cliente.

Familia de productos	Nombre del producto	Versión del producto	Sistema operativo
StorageGRID en NetApp	Almacenamiento de objetos StorageGRID	11,6	n / A
CentOS	CentOS	8,1	CentOS 7.x
Splunk Enterprise	Splunk Enterprise con SmartStore	8.0.3	CentOS 7.x

Requisitos de un solo sitio y de varios sitios

En un entorno de Splunk empresarial (implementaciones medianas y grandes) donde los datos se originan en muchas máquinas y donde muchos usuarios necesitan buscar los datos, puede escalar su implementación distribuyendo instancias de Splunk Enterprise en sitios únicos y múltiples.

Vea los siguientes beneficios de las implementaciones distribuidas de Splunk Enterprise:

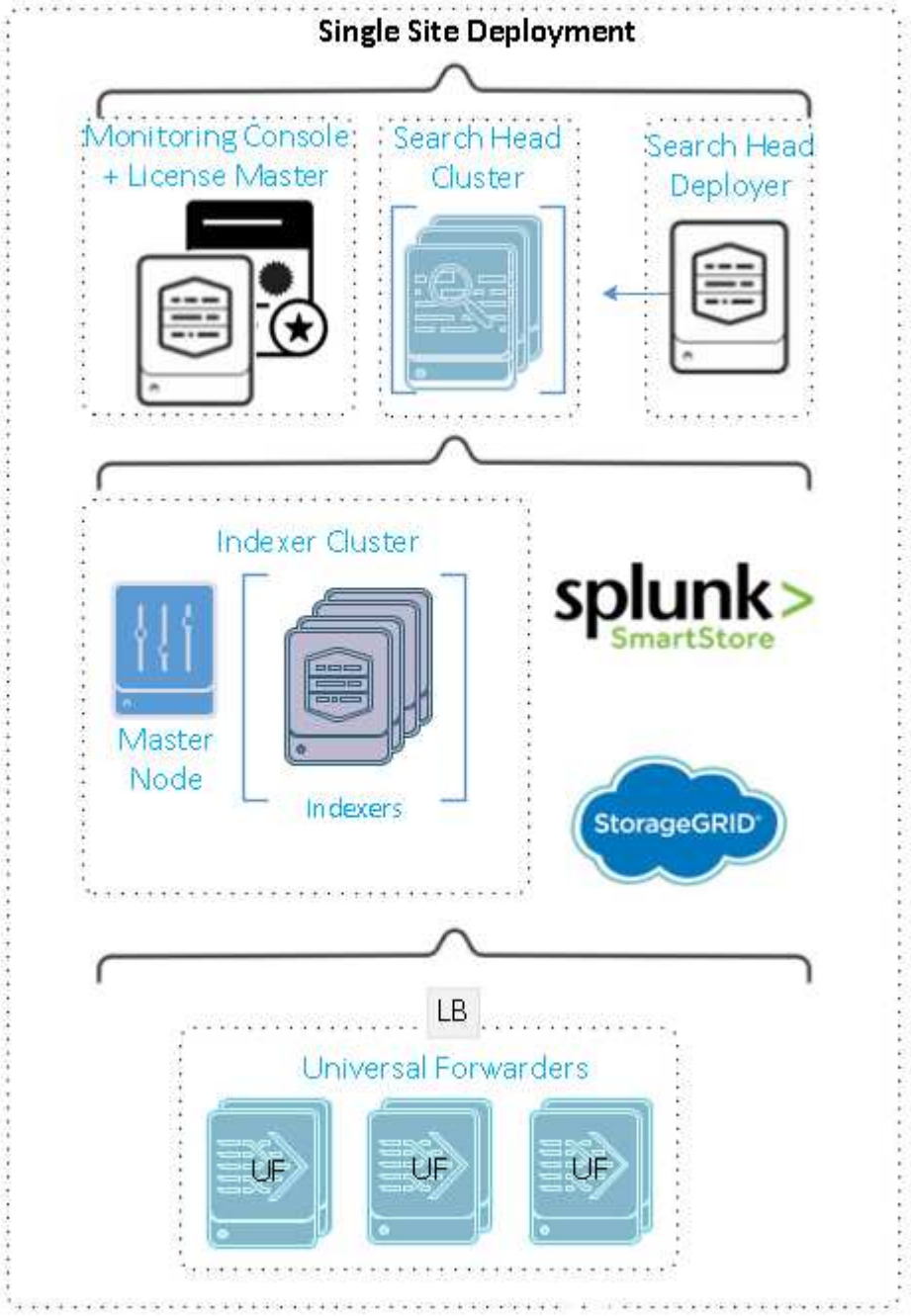
- Acceder a fuentes de datos diversas o dispersas
- Proporcionar funcionalidad para gestionar las necesidades de datos de empresas de cualquier tamaño y complejidad.
- Logre alta disponibilidad y garantice la recuperación ante desastres con replicación de datos e implementación en múltiples sitios

En la siguiente tabla se enumeran los componentes utilizados en un entorno distribuido de Splunk Enterprise.

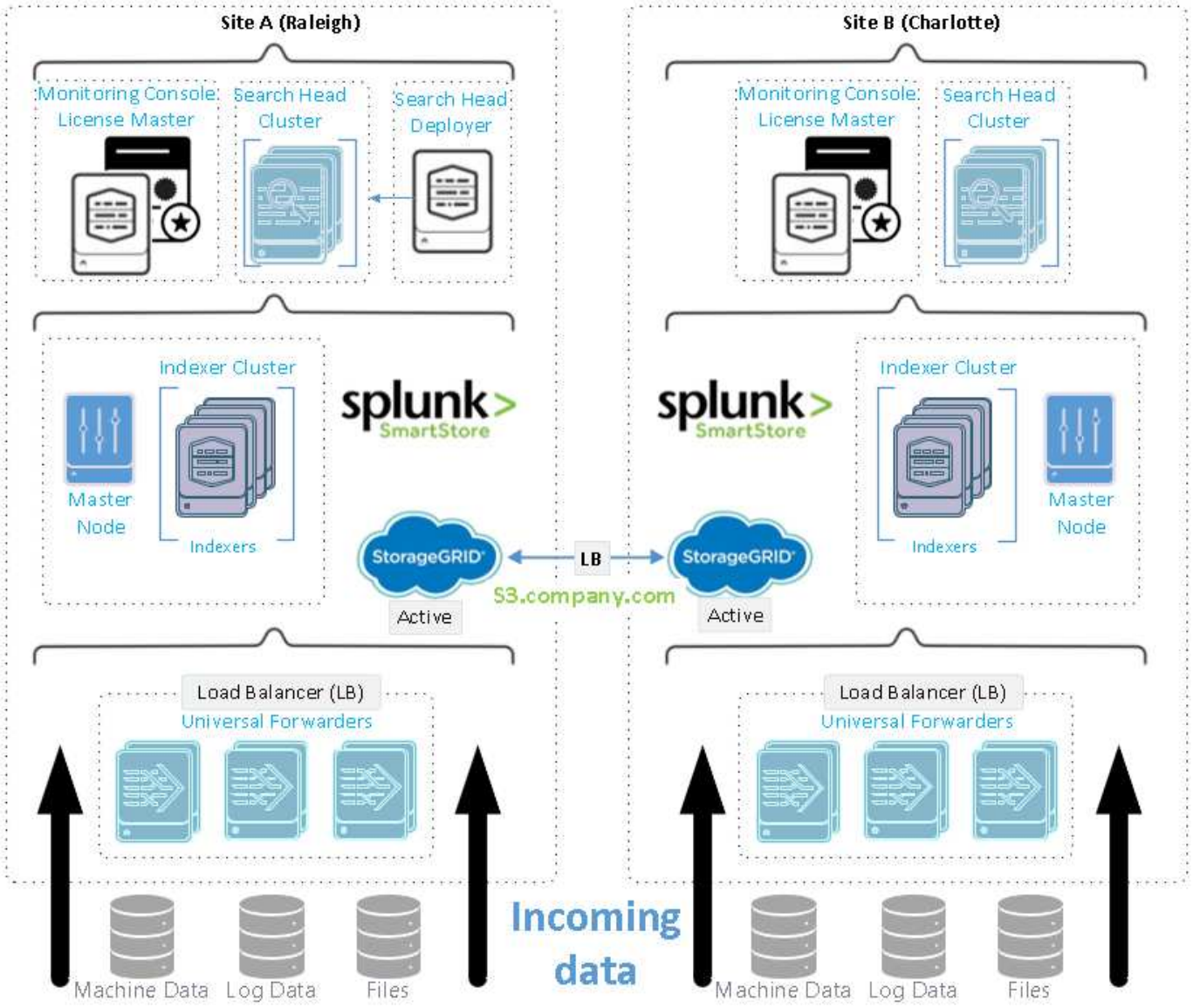
Componente	Descripción	Responsabilidad
Maestro del clúster de índices	Coordina actividades y actualizaciones de un clúster de indexadores	Gestión de índices
Clúster de índices	Grupo de indexadores de Splunk Enterprise que están configurados para replicar los datos de los demás	Indexación
Desplegador de cabezal de búsqueda	Maneja la implementación y las actualizaciones del maestro del clúster.	Gestión de cabezales de búsqueda

Componente	Descripción	Responsabilidad
Grupo de búsqueda de cabezas	Grupo de cabezales de búsqueda que sirve como recurso central para la búsqueda	Gestión de búsquedas
Balanceadores de carga	Lo utilizan los componentes agrupados para gestionar la creciente demanda de los cabezales de búsqueda, los indexadores y el destino S3 para distribuir la carga entre los componentes agrupados.	Gestión de carga para componentes agrupados

Esta figura muestra un ejemplo de una implementación distribuida en un solo sitio.



Esta figura muestra un ejemplo de una implementación distribuida en múltiples sitios.



Requisitos de hardware

Las siguientes tablas enumeran el número mínimo de componentes de hardware necesarios para implementar la solución. Los componentes de hardware que se utilizan en implementaciones específicas de la solución pueden variar según los requisitos del cliente.

> Independientemente de si ha implementado Splunk SmartStore y StorageGRID en un solo sitio o en varios, todos los sistemas se administran desde StorageGRID GRID Manager en un único panel. Consulte la sección "Administración simple con Grid Manager" para obtener más detalles.

Esta tabla enumera el hardware utilizado para un solo sitio.

Hardware	Cantidad	Disco	Capacidad utilizable	Nota
StorageGRID SG1000	1	n / A	n / A	Nodo de administración y balanceador de carga
StorageGRID SG6060	4	x48, 8 TB (disco duro NL-SAS)	1PB	Almacenamiento remoto

Esta tabla enumera el hardware utilizado para una configuración multisitio (por sitio).

Hardware	Cantidad	Disco	Capacidad utilizable	Nota
StorageGRID SG1000	2	n / A	n / A	Nodo de administración y balanceador de carga
StorageGRID SG6060	4	x48, 8 TB (disco duro NL-SAS)	1PB	Almacenamiento remoto

Balanceador de carga NetApp StorageGRID : SG1000

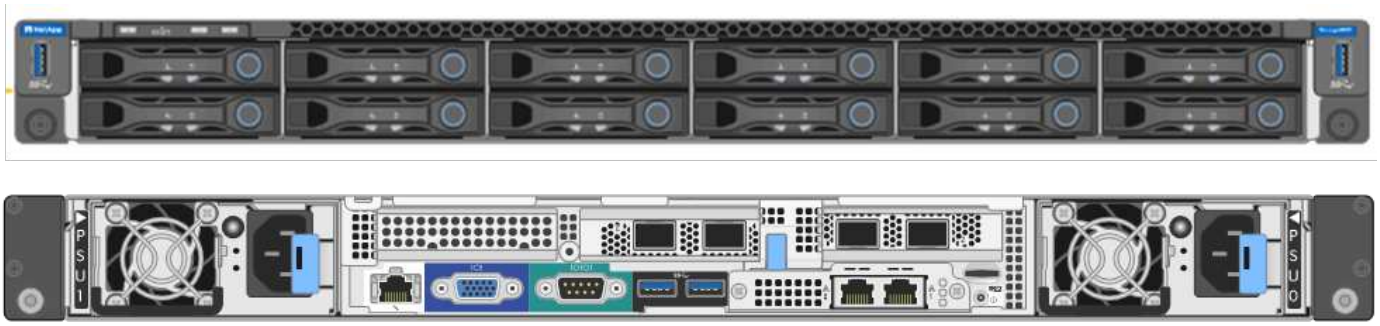
El almacenamiento de objetos requiere el uso de un equilibrador de carga para presentar el espacio de nombres de almacenamiento en la nube. StorageGRID admite balanceadores de carga de terceros de proveedores líderes como F5 y Citrix, pero muchos clientes eligen el balanceador StorageGRID de nivel empresarial por su simplicidad, resiliencia y alto rendimiento. El balanceador de carga StorageGRID está disponible como una máquina virtual, un contenedor o un dispositivo especialmente diseñado.

StorageGRID SG1000 facilita el uso de grupos de alta disponibilidad (HA) y equilibrio de carga inteligente para conexiones de rutas de datos S3. Ningún otro sistema de almacenamiento de objetos local proporciona un equilibrador de carga personalizado.

El aparato SG1000 ofrece las siguientes características:

- Un equilibrador de carga y, opcionalmente, funciones de nodo de administración para un sistema StorageGRID
- El instalador del dispositivo StorageGRID para simplificar la implementación y configuración de nodos
- Configuración simplificada de puntos finales S3 y SSL
- Ancho de banda dedicado (en comparación con compartir un balanceador de carga de terceros con otras aplicaciones)
- Ancho de banda Ethernet agregado de hasta 4 x 100 Gbps

La siguiente imagen muestra el dispositivo SG1000 Gateway Services.

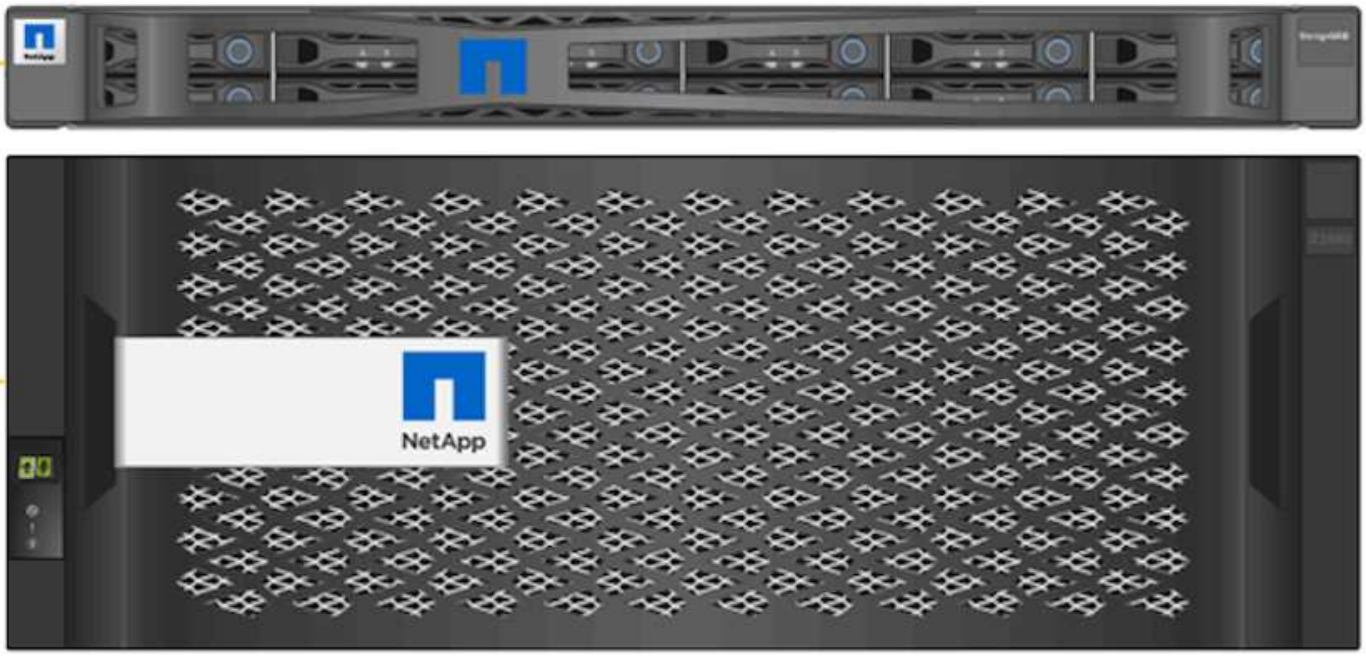


SG6060

El dispositivo StorageGRID SG6060 incluye un controlador de cómputo (SG6060) y un estante de controlador de almacenamiento (E-Series E2860) que contiene dos controladores de almacenamiento y 60 unidades. Este aparato ofrece las siguientes características:

- Escala hasta 400 PB en un solo espacio de nombres.
- Ancho de banda Ethernet agregado de hasta 4 x 25 Gbps.
- Incluye el instalador de dispositivos StorageGRID para simplificar la implementación y configuración de nodos.
- Cada dispositivo SG6060 puede tener uno o dos estantes de expansión adicionales para un total de 180 unidades.
- Dos controladores E-Series E2800 (configuración dúplex) para brindar soporte de conmutación por error del controlador de almacenamiento.
- Estante de unidad de cinco cajones que admite sesenta unidades de 3,5 pulgadas (dos unidades de estado sólido y 58 unidades NL-SAS).

La siguiente imagen muestra el dispositivo SG6060.



Diseño de Splunk

La siguiente tabla enumera la configuración de Splunk para un solo sitio.

Componente de Splunk	Tarea	Cantidad	Núcleos	Memoria	Sistema operativo
Reenvío universal	Responsable de ingerir datos y enviarlos a los indexadores.	4	16 núcleos	32 GB de RAM	CentOS 8.1

Componente de Splunk	Tarea	Cantidad	Núcleos	Memoria	Sistema operativo
Indexador	Gestiona los datos del usuario	10	16 núcleos	32 GB de RAM	CentOS 8.1
Cabezal de búsqueda	La interfaz de usuario busca datos en los indexadores	3	16 núcleos	32 GB de RAM	CentOS 8.1
Desplegador de cabezal de búsqueda	Maneja actualizaciones para grupos de encabezados de búsqueda	1	16 núcleos	32 GB de RAM	CentOS 8.1
Maestro del clúster	Administra la instalación y los indexadores de Splunk.	1	16 núcleos	32 GB de RAM	CentOS 8.1
Consola de monitoreo y maestro de licencias	Realiza una supervisión centralizada de toda la implementación de Splunk y administra las licencias de Splunk.	1	16 núcleos	32 GB de RAM	CentOS 8.1

Las siguientes tablas describen la configuración de Splunk para configuraciones multisitio.

Esta tabla enumera la configuración de Splunk para una configuración multisitio (sitio A).

Componente de Splunk	Tarea	Cantidad	Núcleos	Memoria	Sistema operativo
Reenvío universal	Responsable de ingerir datos y enviarlos a los indexadores.	4	16 núcleos	32 GB de RAM	CentOS 8.1
Indexador	Gestiona los datos del usuario	10	16 núcleos	32 GB de RAM	CentOS 8.1
Cabezal de búsqueda	La interfaz de usuario busca datos en los indexadores	3	16 núcleos	32 GB de RAM	CentOS 8.1
Desplegador de cabezal de búsqueda	Maneja actualizaciones para grupos de encabezados de búsqueda	1	16 núcleos	32 GB de RAM	CentOS 8.1

Componente de Splunk	Tarea	Cantidad	Núcleos	Memoria	Sistema operativo
Maestro del clúster	Administra la instalación y los indexadores de Splunk.	1	16 núcleos	32 GB de RAM	CentOS 8.1
Consola de monitoreo y maestro de licencias	Realiza la supervisión centralizada de toda la implementación de Splunk y administra las licencias de Splunk.	1	16 núcleos	32 GB de RAM	CentOS 8.1

Esta tabla enumera la configuración de Splunk para una configuración multisitio (sitio B).

Componente de Splunk	Tarea	Cantidad	Núcleos	Memoria	Sistema operativo
Reenvío universal	Responsable de ingerir datos y enviarlos a los indexadores.	4	16 núcleos	32 GB de RAM	CentOS 8.1
Indexador	Gestiona los datos del usuario	10	16 núcleos	32 GB de RAM	CentOS 8.1
Cabezal de búsqueda	La interfaz de usuario busca datos en los indexadores	3	16 núcleos	32 GB de RAM	CentOS 8.1
Maestro del clúster	Administra la instalación y los indexadores de Splunk.	1	16 núcleos	32 GB de RAM	CentOS 8.1
Consola de monitoreo y maestro de licencias	Realiza una supervisión centralizada de toda la implementación de Splunk y administra las licencias de Splunk.	1	16 núcleos	32 GB de RAM	CentOS 8.1

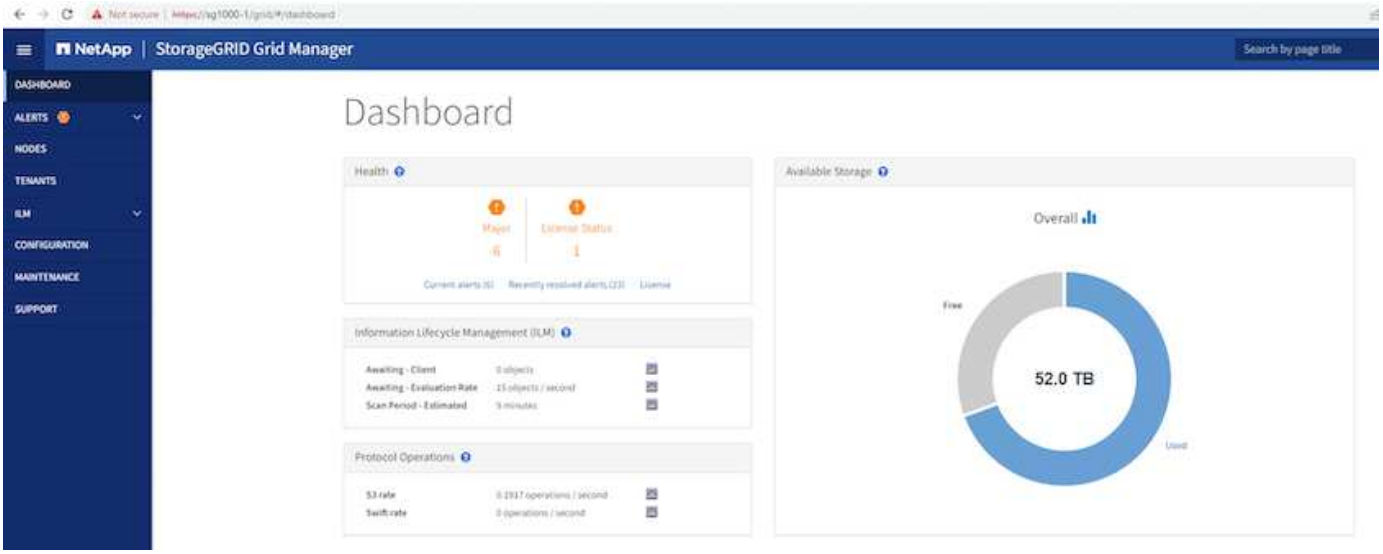
Funciones flexibles de StorageGRID para Splunk SmartStore

StorageGRID tiene una amplia variedad de características que los usuarios pueden

aprovechar y personalizar para su entorno en constante cambio. Desde la implementación hasta la ampliación de su Splunk SmartStore, su entorno exige una rápida adopción de los cambios y no debe interrumpir el funcionamiento de Splunk. Las políticas de administración de datos flexibles (ILM) y los clasificadores de tráfico (QoS) de StorageGRID le permiten planificar y adaptarse a su entorno.

Gestión sencilla con Grid Manager

Grid Manager es la interfaz gráfica basada en navegador que le permite configurar, administrar y monitorear su sistema StorageGRID en ubicaciones distribuidas globalmente en un solo panel, como se muestra en la siguiente imagen.



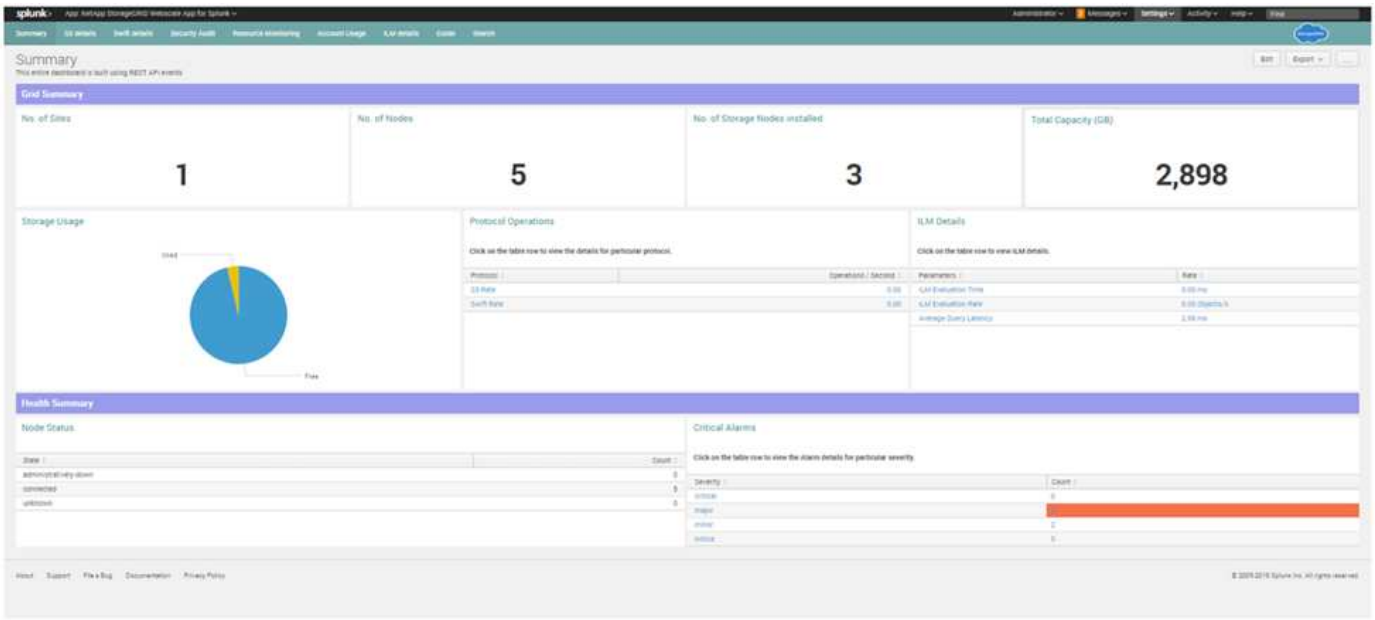
Realice las siguientes tareas con la interfaz de Grid Manager:

- Administre repositorios de objetos, como imágenes, vídeos y registros, distribuidos globalmente y a escala de petabytes.
- Supervisar los nodos y servicios de la red para garantizar la disponibilidad de los objetos.
- Gestione la ubicación de los datos de objetos a lo largo del tiempo utilizando reglas de gestión del ciclo de vida de la información (ILM). Estas reglas rigen lo que sucede con los datos de un objeto después de su ingesta, cómo se protegen contra pérdidas, dónde se almacenan los datos del objeto y durante cuánto tiempo.
- Supervisar transacciones, rendimiento y operaciones dentro del sistema.

Aplicación NetApp StorageGRID para Splunk

La aplicación NetApp StorageGRID para Splunk es una aplicación específica para Splunk Enterprise. Esta aplicación funciona junto con el complemento NetApp StorageGRID para Splunk. Proporciona visibilidad sobre el estado de StorageGRID , información de uso de la cuenta, detalles de auditoría de seguridad, uso y monitoreo de recursos, etc.

La siguiente imagen muestra la aplicación StorageGRID para Splunk.



Políticas de ILM

StorageGRID tiene políticas de administración de datos flexibles que incluyen mantener múltiples copias de sus objetos y usar esquemas EC (codificación de borrado) como 2+1 y 4+2 (y muchos otros) para almacenar sus objetos dependiendo de los requisitos específicos de rendimiento y protección de datos. A medida que las cargas de trabajo y los requisitos cambian con el tiempo, es común que las políticas de ILM también deban cambiar con el tiempo. La modificación de las políticas de ILM es una característica fundamental que permite a los clientes de StorageGRID adaptarse a su entorno en constante cambio de forma rápida y sencilla.

Actuación

StorageGRID escala el rendimiento agregando más nodos, que pueden ser máquinas virtuales, hardware o dispositivos especialmente diseñados como SG5712, SG5760, SG6060 o SGF6024. En nuestras pruebas, superamos los requisitos clave de rendimiento de SmartStore con una red de tres nodos de tamaño mínimo utilizando el dispositivo SG6060. A medida que los clientes amplían su infraestructura Splunk con indexadores adicionales, pueden agregar más nodos de almacenamiento para aumentar el rendimiento y la capacidad.

Configuración del balanceador de carga y del punto final

Los nodos de administración en StorageGRID proporcionan la interfaz de usuario (IU) de Grid Manager y el punto final de API REST para ver, configurar y administrar su sistema StorageGRID , así como registros de auditoría para rastrear la actividad del sistema. Para proporcionar un punto final S3 de alta disponibilidad para el almacenamiento remoto Splunk SmartStore, implementamos el balanceador de carga StorageGRID , que se ejecuta como un servicio en los nodos de administración y los nodos de puerta de enlace. Además, el balanceador de carga también administra el tráfico local y se comunica con GSLB (Global Server Load Balancing) para ayudar con la recuperación ante desastres.

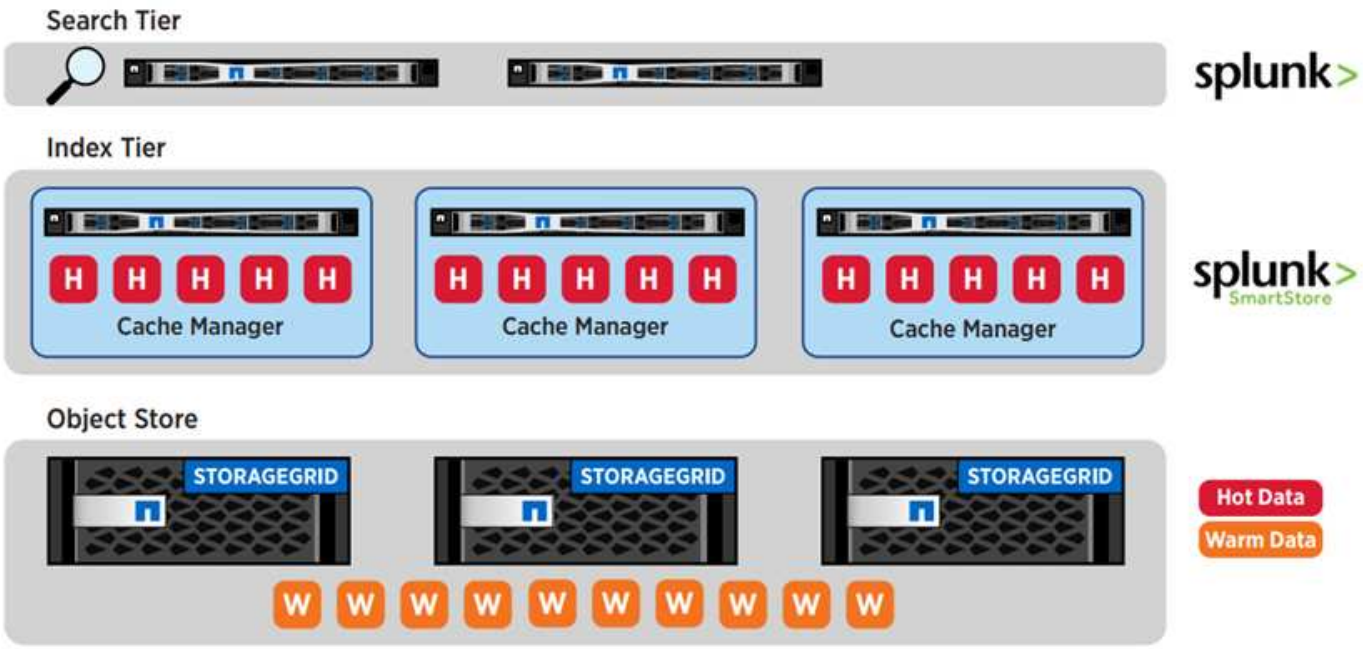
Para mejorar aún más la configuración de los puntos finales, StorageGRID proporciona políticas de clasificación de tráfico integradas en el nodo de administración, le permite monitorear el tráfico de su carga de trabajo y aplicar varios límites de calidad de servicio (QoS) a sus cargas de trabajo. Las políticas de clasificación de tráfico se aplican a los puntos finales del servicio StorageGRID Load Balancer para los nodos de puerta de enlace y los nodos de administración. Estas políticas pueden ayudar a limitar y monitorear el tráfico.

Nivelación inteligente y ahorro de costes

A medida que los clientes se dan cuenta del poder y la facilidad de usar el análisis de datos de Splunk, naturalmente quieren indexar una cantidad cada vez mayor de datos. A medida que crece la cantidad de datos, también crece la infraestructura de procesamiento y almacenamiento necesaria para brindarles servicio. Como los datos más antiguos se consultan con menos frecuencia, comprometer la misma cantidad de recursos computacionales y consumir el costoso almacenamiento primario se vuelve cada vez más ineficiente. Para operar a escala, los clientes se benefician al mover datos calientes a un nivel más rentable, liberando capacidad de procesamiento y almacenamiento primario para datos calientes.

Splunk SmartStore con StorageGRID ofrece a las organizaciones una solución escalable, de alto rendimiento y rentable. Debido a que SmartStore reconoce los datos, evalúa automáticamente los patrones de acceso a los datos para determinar qué datos deben ser accesibles para análisis en tiempo real (datos activos) y qué datos deben residir en un almacenamiento a largo plazo de menor costo (datos tibios). SmartStore utiliza la API AWS S3 estándar de la industria de forma dinámica e inteligente, colocando los datos en el almacenamiento S3 proporcionado por StorageGRID. La arquitectura de escalamiento flexible de StorageGRID permite que el nivel de datos cálidos crezca de manera rentable según sea necesario. La arquitectura basada en nodos de StorageGRID garantiza que los requisitos de rendimiento y costos se cumplan de manera óptima.

La siguiente imagen ilustra la organización en niveles de Splunk y StorageGRID .

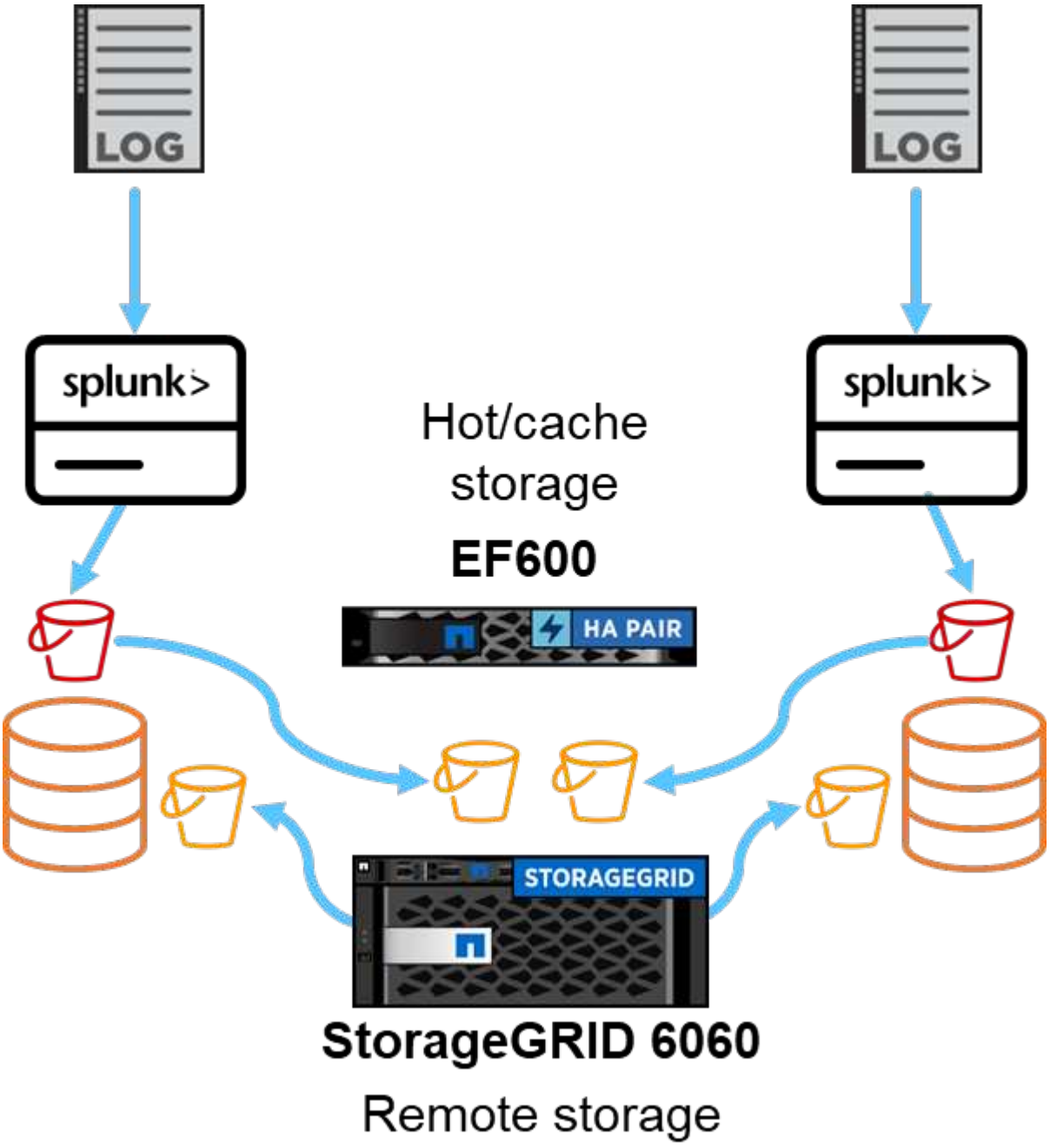


La combinación líder en la industria de Splunk SmartStore con NetApp StorageGRID ofrece los beneficios de la arquitectura desacoplada a través de una solución de pila completa.

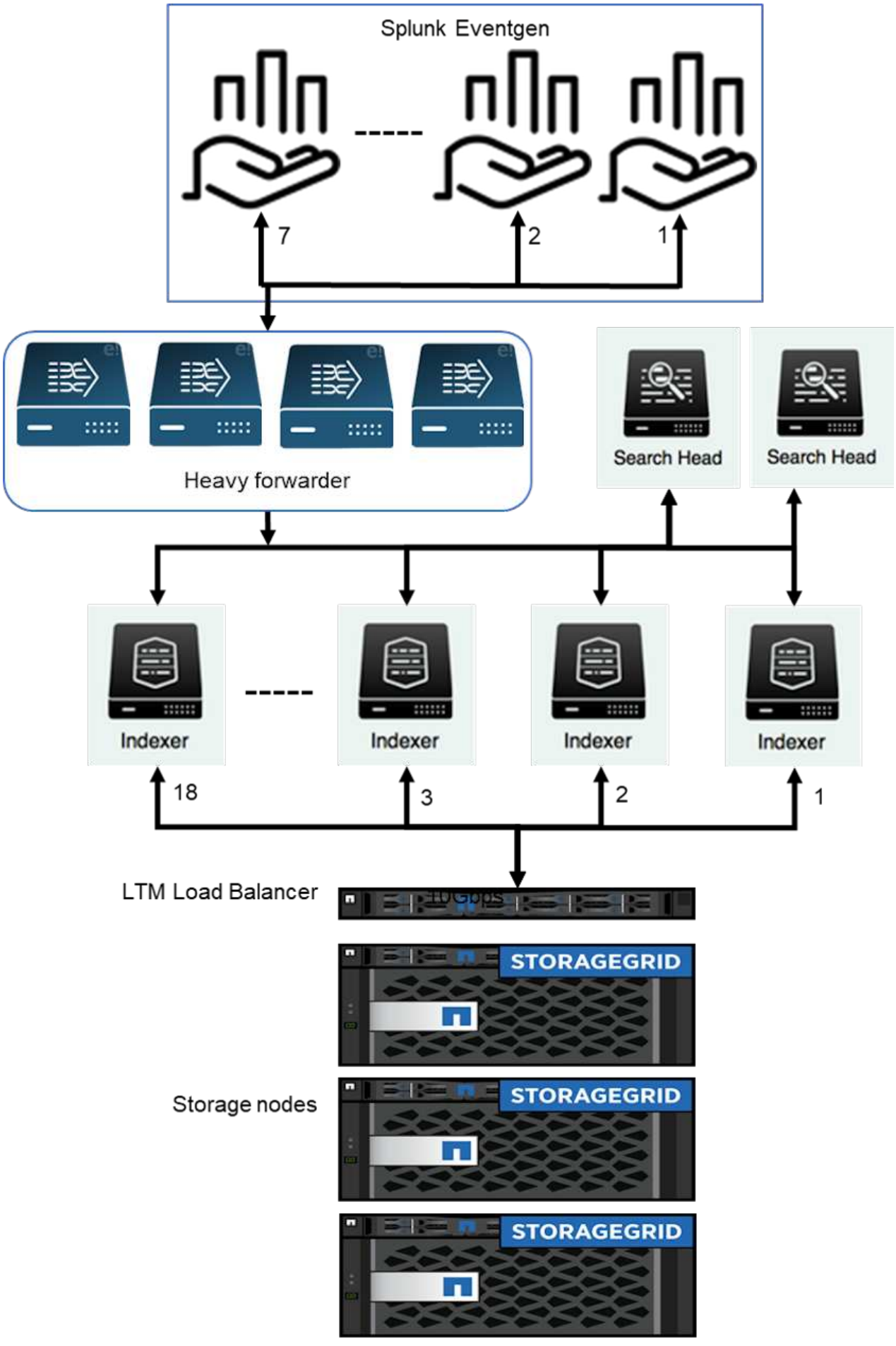
Rendimiento de SmartStore en un solo sitio

Esta sección describe el rendimiento de Splunk SmartStore en un controlador NetApp StorageGRID . Splunk SmartStore mueve datos cálidos al almacenamiento remoto, que

en este caso es el almacenamiento de objetos StorageGRID en la validación del rendimiento.



Utilizamos EF600 para almacenamiento en caché/activo y StorageGRID 6060 para almacenamiento remoto. Utilizamos la siguiente arquitectura para la validación del rendimiento. Utilizamos dos cabezales de búsqueda, cuatro reenvíos pesados para enviar los datos a los indexadores, siete generadores de eventos Splunk (Eventgens) para generar los datos en tiempo real y 18 indexadores para almacenar los datos.


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

Configuración

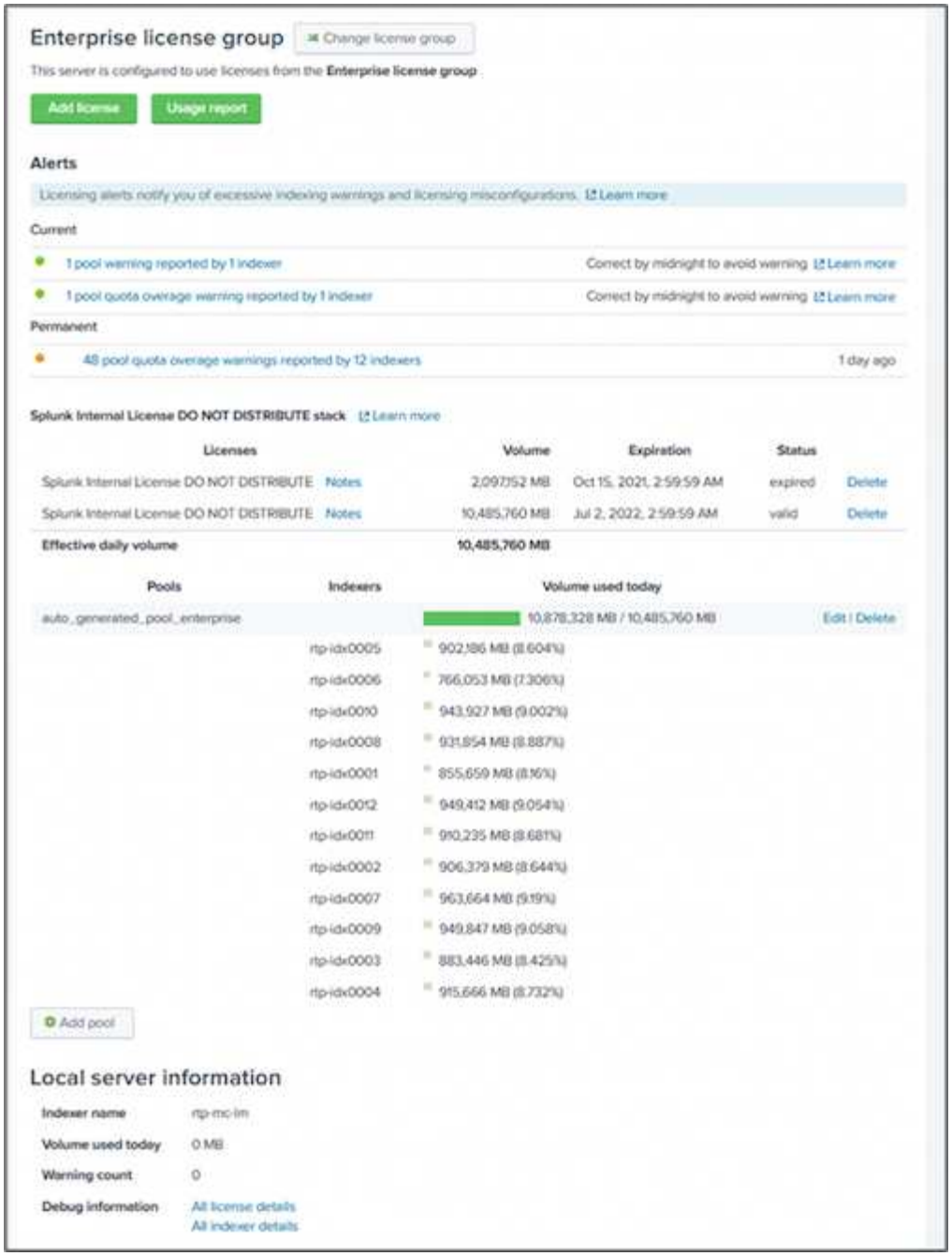
Esta tabla enumera el hardware utilizado para la validación del rendimiento de SmartStorage.

Componente de Splunk	Tarea	Cantidad	Núcleos	Memoria	Sistema operativo
Transportador pesado	Responsable de ingerir datos y enviarlos a los indexadores.	4	16 núcleos	32 GB de RAM	TRINEO 15 SP2
Indexador	Gestiona los datos del usuario	18	16 núcleos	32 GB de RAM	TRINEO 15 SP2
Cabezal de búsqueda	El frontend del usuario busca datos en los indexadores	2	16 núcleos	32 GB de RAM	TRINEO 15 SP2
Desplegador de cabezal de búsqueda	Maneja actualizaciones para grupos de encabezados de búsqueda	1	16 núcleos	32 GB de RAM	TRINEO 15 SP2
Maestro del clúster	Administra la instalación y los indexadores de Splunk.	1	16 núcleos	32 GB de RAM	TRINEO 15 SP2
Consola de monitoreo y maestro de licencias	Realiza una supervisión centralizada de toda la implementación de Splunk y administra las licencias de Splunk.	1	16 núcleos	32 GB de RAM	TRINEO 15 SP2

Validación del rendimiento de la tienda remota SmartStore

En esta validación de rendimiento, configuramos el caché SmartStore en el almacenamiento local en todos los indexadores para 10 días de datos. Hemos habilitado el `maxDataSize=auto` (tamaño de depósito de 750 MB) en el administrador de clústeres de Splunk y envió los cambios a todos los indexadores. Para medir el rendimiento de carga, ingerimos 10 TB por día durante 10 días y transferimos todos los buckets activos a cálidos al mismo tiempo y capturamos el rendimiento máximo y promedio por instancia y en toda la implementación desde el panel de control de la consola de monitoreo de SmartStore.

Esta imagen muestra los datos ingeridos en un día.



Ejecutamos el siguiente comando desde el clúster maestro (el nombre del índice es `eventgen-test`). Luego, capturamos el rendimiento de carga máximo y promedio por instancia y en toda la implementación a través de los paneles de control de la consola de monitoreo de SmartStore.

```
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
```

El maestro del clúster tiene autenticación sin contraseña para todos los indexadores (rtp-idx0001…rtp-idx0018).

Para medir el rendimiento de la descarga, eliminamos todos los datos de la memoria caché ejecutando la CLI de desalojo dos veces usando el siguiente comando.

Ejecutamos el siguiente comando desde el clúster maestro y ejecutamos la búsqueda desde el cabezal de búsqueda sobre 10 días de datos del almacén remoto de StorageGRID. Luego capturamos el rendimiento de carga máximo y promedio por instancia y en toda la implementación a través de los paneles de control de la consola de monitoreo de SmartStore.

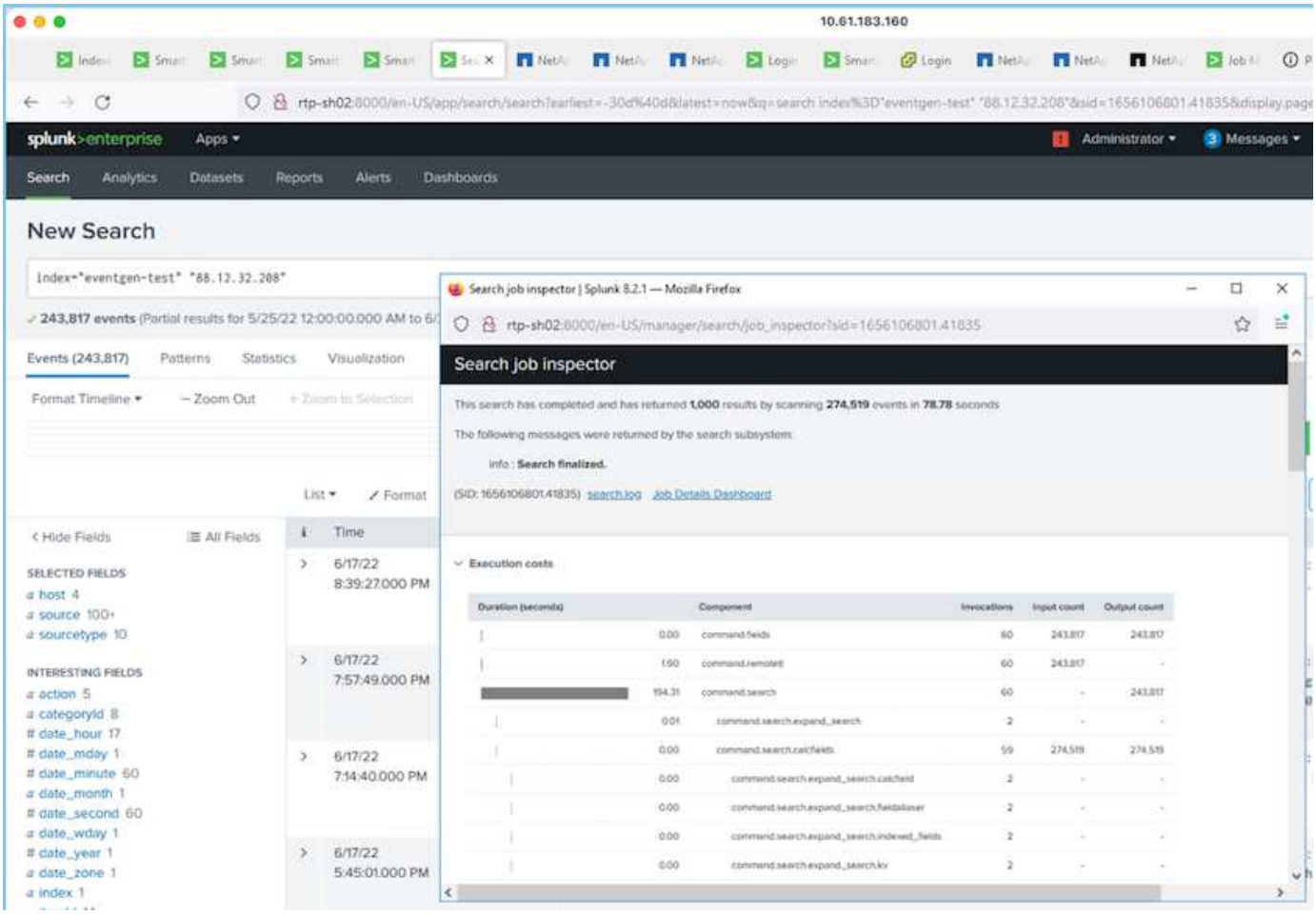
```
for i in rtp-idx0001 rtp-idx0002 rtp-idx0003 rtp-idx0004 rtp-idx0005 rtp-
idx0006 rtp-idx0007 rtp-idx0008 rtp-idx0009 rtp-idx0010 rtp-idx0011 rtp-
idx0012 rtp-idx0013 rtp-idx0014 rtp-idx0015 rtp-idx0016 rtp-idx0017 rtp-
idx0018 ; do  ssh $i " hostname;  date; /opt/splunk/bin/splunk _internal
call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

Las configuraciones del indexador se enviaron desde el maestro del clúster SmartStore. El maestro del clúster tenía la siguiente configuración para el indexador.

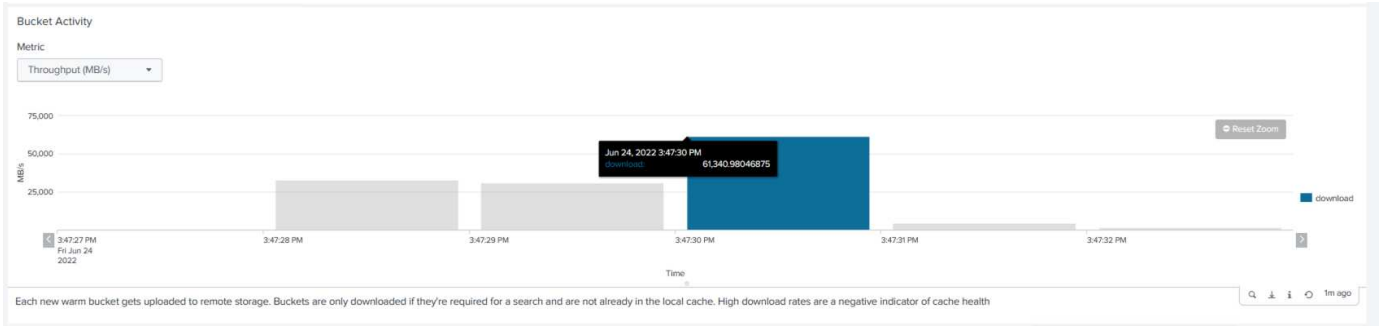
```
Rtp-cm01:~ # cat /opt/splunk/etc/master-apps/_cluster/local/indexes.conf
[default]
maxDataSize = auto
#defaultDatabase = eventgen-basic
defaultDatabase = eventgen-test
hotlist_recency_secs = 864000
repFactor = auto
[volume:remote_store]
storageType = remote
path = s3://smartstore2
remote.s3.access_key = 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:~ #
```

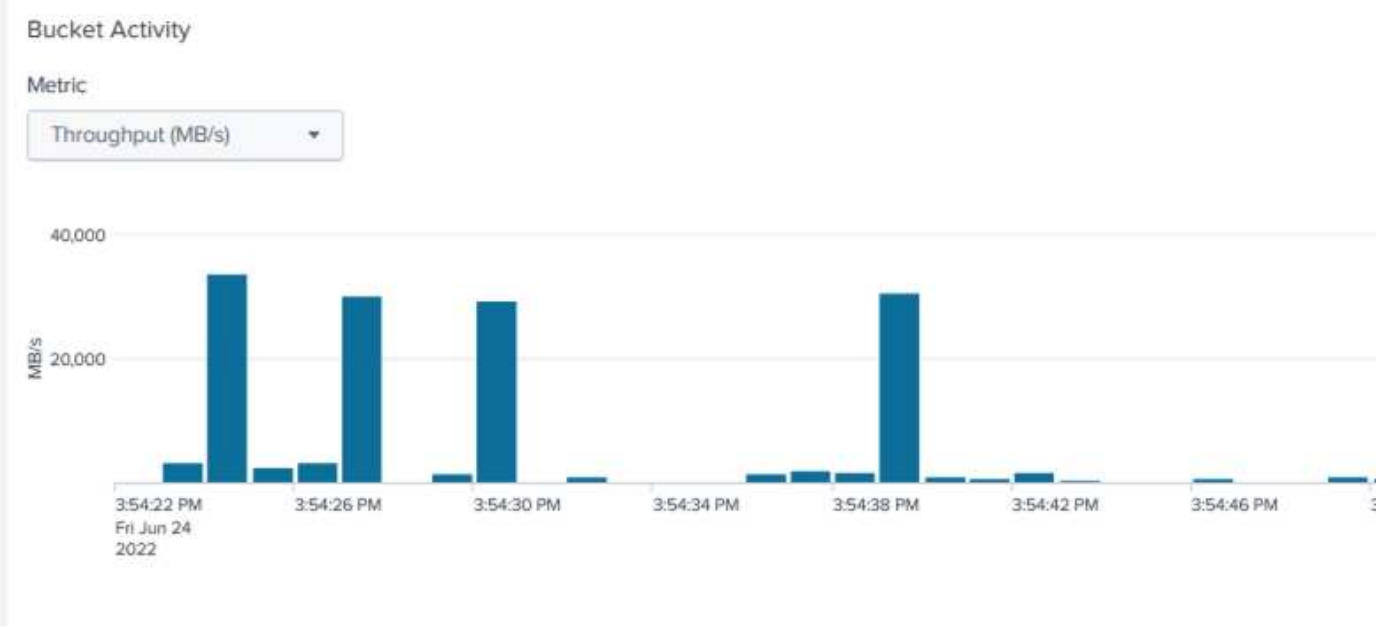
Ejecutamos la siguiente consulta de búsqueda en el encabezado de búsqueda para recopilar la matriz de rendimiento.



Recopilamos la información de rendimiento del clúster maestro. El rendimiento máximo fue de 61,34 GBps.



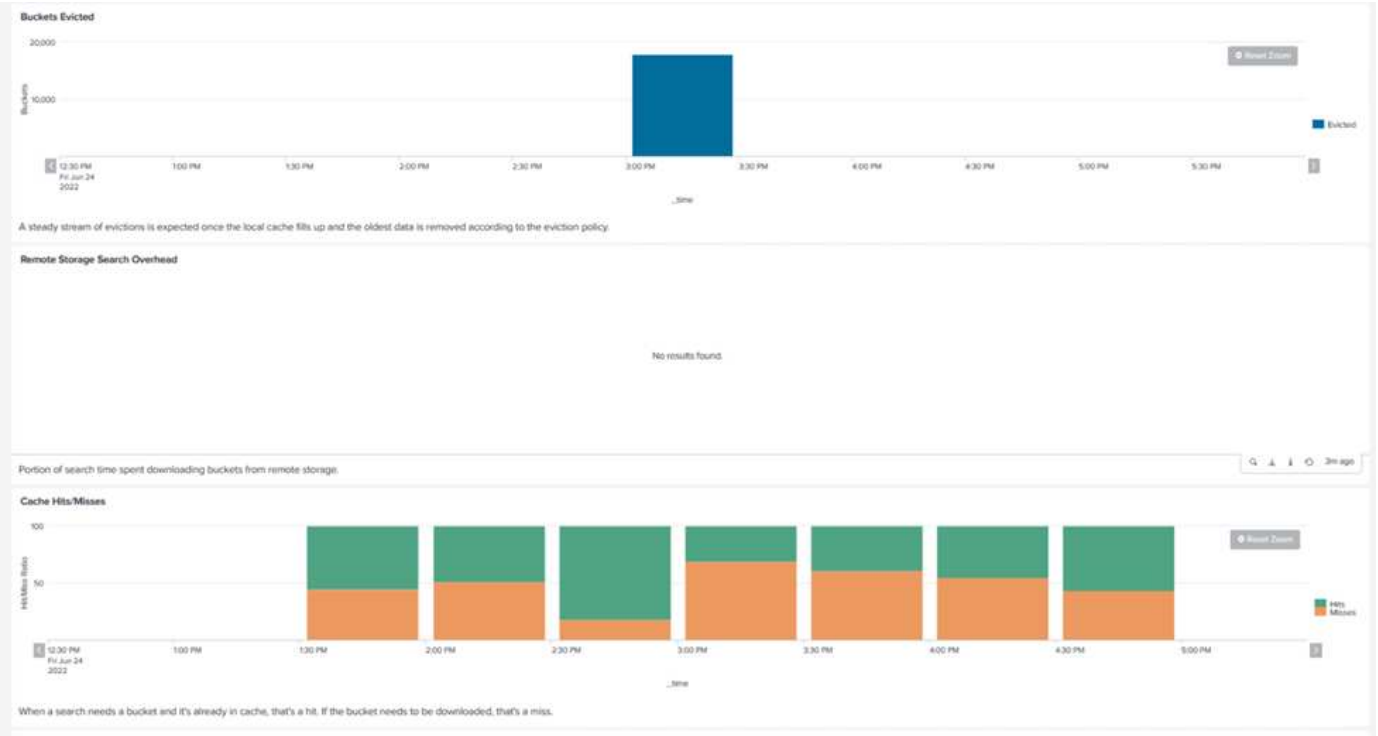
El rendimiento promedio fue de aproximadamente 29 GBps.



Rendimiento de StorageGRID

El rendimiento de SmartStore se basa en la búsqueda de patrones y cadenas específicos entre grandes cantidades de datos. En esta validación, los eventos se generan utilizando "Eventgen" en un índice de Splunk específico (eventgen-test) a través del cabezal de búsqueda y la solicitud se dirige a StorageGRID para la mayoría de las consultas. La siguiente imagen muestra los aciertos y errores de los datos de la consulta. Los datos de aciertos provienen del disco local y los datos de errores provienen del controlador StorageGRID .

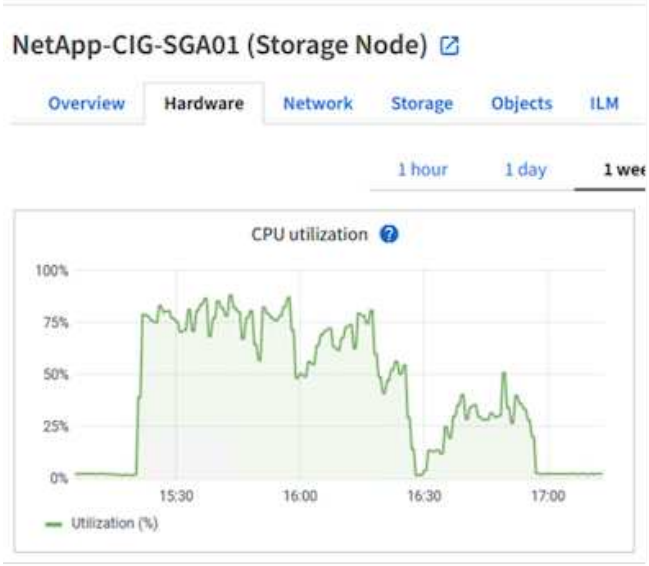
El color verde muestra los datos de aciertos y el color naranja muestra los datos de errores.

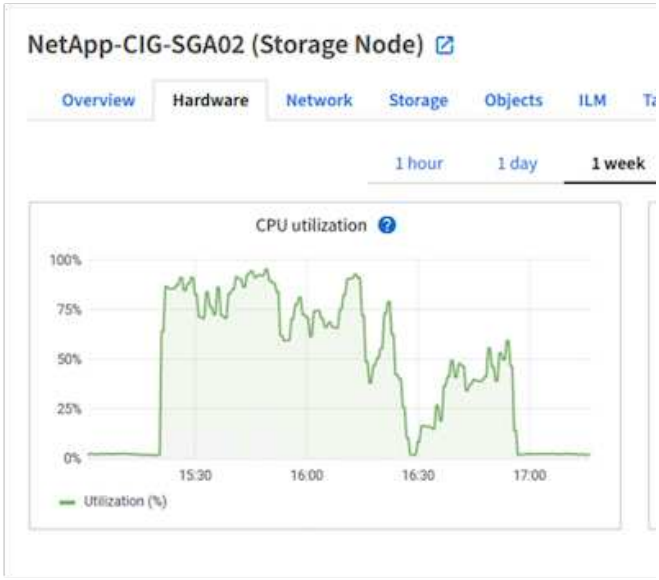


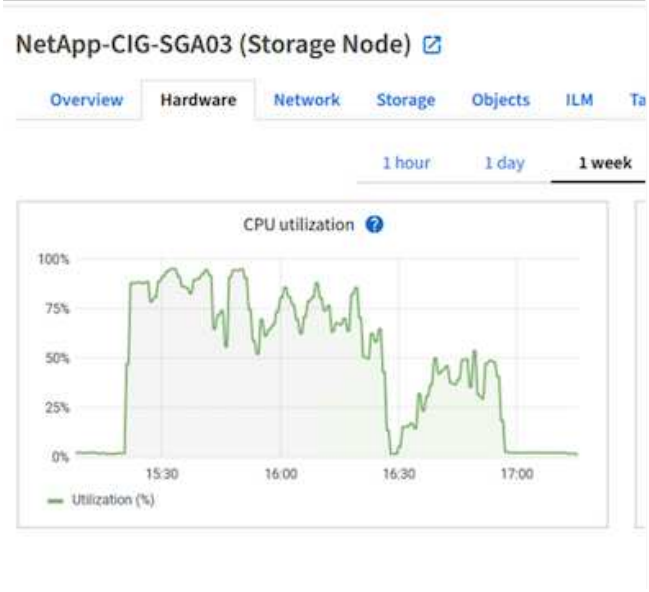
Cuando se ejecuta la consulta para la búsqueda en StorageGRID, el tiempo de recuperación de S3 de StorageGRID se muestra en la siguiente imagen.

Uso del hardware de StorageGRID

La instancia de StorageGRID tiene un equilibrador de carga y tres controladores StorageGRID . La utilización de la CPU para los tres controladores es del 75% al 100%.







SmartStore con controlador de almacenamiento NetApp : beneficios para el cliente

- **Desacoplamiento entre computación y almacenamiento.** Splunk SmartStore desacopla el procesamiento y el almacenamiento, lo que le ayuda a escalarlos de forma independiente.
- **Datos bajo demanda.** SmartStore acerca los datos al procesamiento a pedido y brinda elasticidad de procesamiento y almacenamiento y eficiencia de costos para lograr una retención de datos más prolongada a escala.
- **Compatible con API AWS S3.** SmartStore utiliza la API de AWS S3 para comunicarse con el almacenamiento de restauración, que es un almacén de objetos compatible con AWS S3 y la API de S3, como StorageGRID.
- **Reduce los requisitos y costos de almacenamiento.** SmartStore reduce los requisitos de almacenamiento para datos antiguos (cálidos/fríos). Solo necesita una única copia de datos porque el almacenamiento de NetApp brinda protección de datos y se encarga de las fallas y la alta disponibilidad.
- **Fallo de hardware.** La falla de un nodo en una implementación de SmartStore no hace que los datos sean inaccesibles y tiene una recuperación del indexador mucho más rápida en caso de falla de hardware o desequilibrio de datos.

- Caché consciente de aplicaciones y datos.
- Agregar y quitar indexadores y configurar y desmontar clústeres a pedido.
- El nivel de almacenamiento ya no está vinculado al hardware.

Conclusión

Splunk Enterprise es la solución SIEM líder en el mercado que impulsa resultados en los equipos de seguridad, TI y DevOps. El uso de Splunk ha aumentado considerablemente en las organizaciones de nuestros clientes. Por lo tanto, es necesario agregar más fuentes de datos y, al mismo tiempo, conservar los datos durante un período más largo, lo que tensiona la infraestructura de Splunk.

La combinación de Splunk SmartStore y NetApp StorageGRID está diseñada para proporcionar una arquitectura escalable para que las organizaciones logren un mejor rendimiento de ingesta con el almacenamiento de objetos SmartStore y StorageGRID y una mayor escalabilidad para un entorno Splunk en múltiples regiones geográficas.

Dónde encontrar información adicional

Para obtener más información sobre la información que se describe en este documento, revise los siguientes documentos y/o sitios web:

- "Recursos de documentación de NetApp StorageGRID"
- "Documentación de productos de NetApp"
- "Documentación de Splunk Enterprise"
- "Splunk Enterprise Acerca de SmartStore"
- "Manual de implementación distribuida de Splunk Enterprise"
- "Splunk Enterprise: administración de indexadores y clústeres de indexadores"

Información de copyright

Información de la marca comercial