NetApp StorageGRID mit Splunk SmartStore

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Inhalt

NetApp StorageGRID mit Splunk SmartStore

TR-4869: NetApp StorageGRID mit Splunk SmartStore

Splunk Enterprise ist die marktführende SIEM-Lösung (Security Information and Event Management), die in den Sicherheits-, IT- und DevOps-Teams zu Ergebnissen führt.

Überblick

Das Datenvolumen wächst weiterhin exponentiell und schafft enorme Chancen für Unternehmen, die diese enorme Ressource nutzen können. Splunk Enterprise wird in immer mehr Anwendungsfällen immer häufiger eingesetzt. Mit der Zunahme der Anwendungsfälle wächst auch die Datenmenge, die Splunk Enterprise aufnimmt und verarbeitet. Die traditionelle Architektur von Splunk Enterprise ist ein verteiltes Scale-Out-Design, das hervorragenden Datenzugriff und hervorragende Datenverfügbarkeit bietet. Unternehmen, die diese Architektur verwenden, sind jedoch mit steigenden Kosten konfrontiert, die mit der Skalierung verbunden sind, um das schnell wachsende Datenvolumen zu bewältigen.

Splunk SmartStore mit NetApp StorageGRID löst diese Herausforderung durch die Bereitstellung eines neuen Bereitstellungsmodells, bei dem Rechenleistung und Speicher entkoppelt sind. Diese Lösung ermöglicht außerdem eine unübertroffene Skalierbarkeit und Elastizität für Splunk Enterprise-Umgebungen, indem sie Kunden die Skalierung über einzelne und mehrere Standorte hinweg ermöglicht. Gleichzeitig werden die Kosten gesenkt, indem Rechenleistung und Speicher unabhängig voneinander skaliert werden und dem kostengünstigen, Cloud-basierten S3-Objektspeicher intelligentes Tiering hinzugefügt wird.

Die Lösung optimiert die Datenmenge im lokalen Speicher, während die Suchleistung erhalten bleibt, sodass Rechenleistung und Speicher nach Bedarf skaliert werden können. SmartStore wertet automatisch Datenzugriffsmuster aus, um zu bestimmen, welche Daten für Echtzeitanalysen zugänglich sein müssen und welche Daten im kostengünstigeren S3-Objektspeicher gespeichert werden sollten.

Dieser technische Bericht beschreibt den Nutzen, den NetApp einer Splunk SmartStore-Lösung bietet, und demonstriert gleichzeitig ein Framework für die Gestaltung und Dimensionierung von Splunk SmartStore in Ihrer Umgebung. Das Ergebnis ist eine einfache, skalierbare und belastbare Lösung mit überzeugenden Gesamtbetriebskosten. StorageGRID bietet den skalierbaren und kostengünstigen Objektspeicher auf Basis des S3-Protokolls/API, auch als Remote-Speicher bekannt, sodass Unternehmen ihre Splunk-Lösung kostengünstiger skalieren und gleichzeitig die Ausfallsicherheit erhöhen können.

Splunk SmartStore bezeichnet Objektspeicher als Remote-Speicher oder Remote-Speicherebenen.

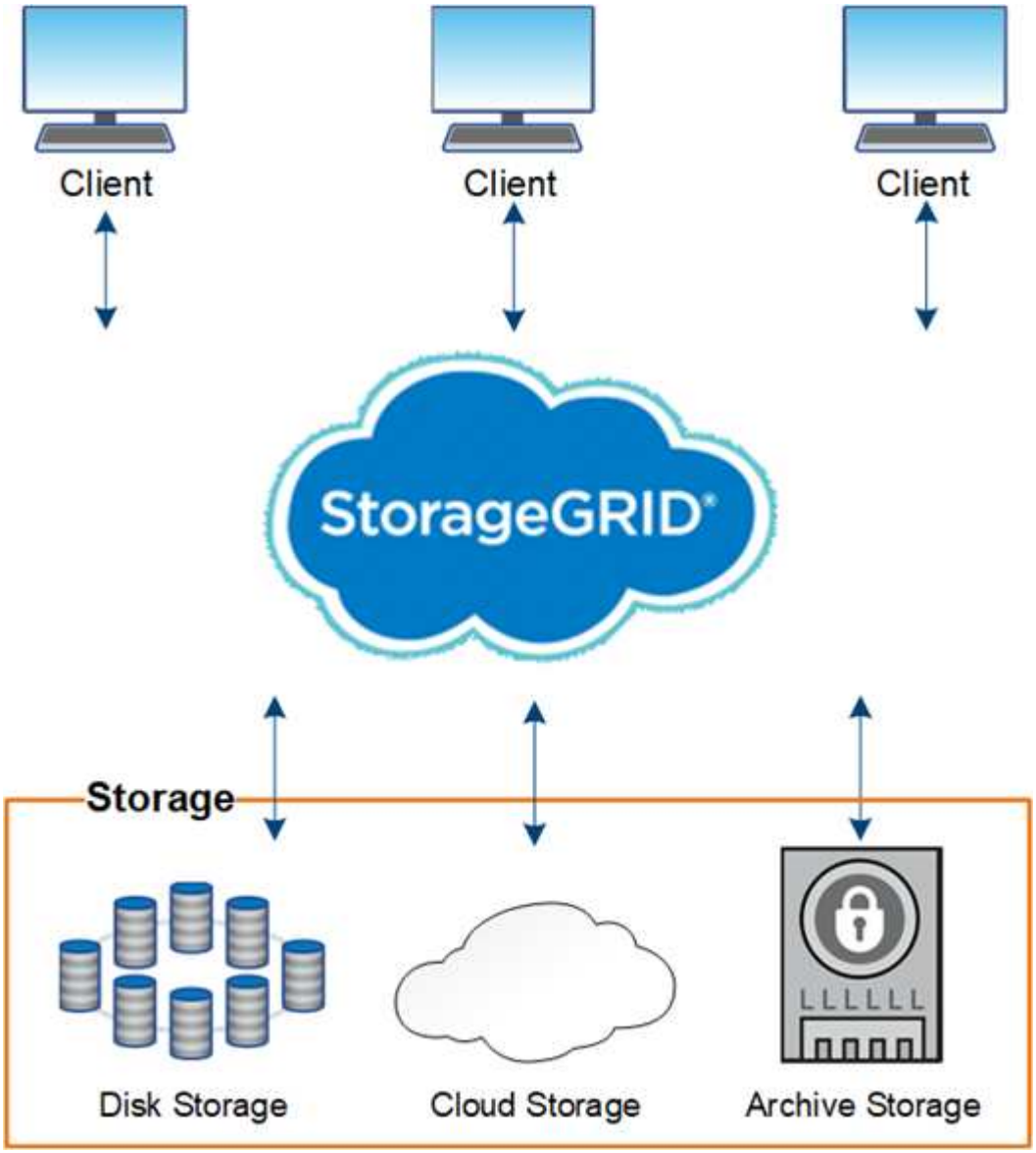
Über NetApp StorageGRID

NetApp StorageGRID ist eine softwaredefinierte Objektspeicherlösung für große Archive, Medienrepositorys und Webdatenspeicher. Mit StorageGRID nutzt NetApp zwei Jahrzehnte Erfahrung in der Bereitstellung branchenführender Innovations- und Datenmanagementlösungen und unterstützt Unternehmen dabei, den Wert ihrer Informationen sowohl vor Ort als auch in öffentlichen, privaten oder hybriden Cloud-Bereitstellungen zu verwalten und zu maximieren.

StorageGRID bietet sicheren, dauerhaften Speicher für unstrukturierte Daten in großem Umfang. Integrierte, metadatengesteuerte Richtlinien zur Lebenszyklusverwaltung optimieren den Verbleib Ihrer Daten während ihrer gesamten Lebensdauer. Um die Kosten zu senken, werden Inhalte zur richtigen Zeit am richtigen Ort und auf der richtigen Speicherebene platziert. Der einzelne Namespace ermöglicht den Zugriff auf die Daten über einen einzigen Aufruf, unabhängig vom geografischen Standort des StorageGRID Speichers. Kunden können

mehrere StorageGRID Instanzen zwischen Rechenzentren und in der Cloud-Infrastruktur bereitstellen und verwalten.

Ein StorageGRID -System besteht aus global verteilten, redundanten, heterogenen Knoten, die sowohl in bestehende als auch in Client-Anwendungen der nächsten Generation integriert werden können.



IDC MarketScape hat NetApp kürzlich im neuesten Bericht „IDC MarketScape: Worldwide Object-Based Storage 2019 Vendor Assessment“ als führend bezeichnet. Mit fast 20 Jahren Erfahrung im Produktionseinsatz in den anspruchsvollsten Branchen ist StorageGRID ein anerkannter Marktführer im Bereich unstrukturierter Daten.

Mit StorageGRID können Sie Folgendes erreichen:

- Stellen Sie mehrere StorageGRID -Instanzen bereit, um über einen einzigen Namespace, der problemlos auf Hunderte von Petabyte skaliert werden kann, von jedem Standort zwischen Rechenzentren und der Cloud auf Daten zuzugreifen.
- Bieten Sie Flexibilität bei der Bereitstellung und zentralen Verwaltung über Infrastrukturen hinweg.
- Sorgen Sie für unübertroffene Haltbarkeit mit einer Haltbarkeit von 99,9999 % durch Nutzung von mehrschichtigem Erasure Coding (EC).

- Aktivieren Sie mehr Hybrid-Multi-Cloud-Funktionen mit validierten Integrationen in Amazon S3 Glacier und Azure Blob.
- Erfüllen Sie gesetzliche Verpflichtungen und erleichtern Sie die Einhaltung durch manipulationssichere Datenspeicherung, ohne proprietäre APIs oder Anbieterabhängigkeit.

Weitere Informationen dazu, wie StorageGRID Ihnen bei der Lösung Ihrer komplexesten Probleme im Bereich der unstrukturierten Datenverwaltung helfen kann, finden Sie im "NetApp StorageGRID Homepage" .

Über Splunk Enterprise

Splunk Enterprise ist eine Plattform, mit der Daten in Taten umgesetzt werden. Von verschiedenen Quellen wie Protokolldateien, Websites, Geräten, Sensoren und Anwendungen generierte Daten werden an die Splunk-Indexer gesendet und von ihnen analysiert, sodass Sie aus den Daten umfassende Erkenntnisse gewinnen können. Es kann Datenlecks aufdecken, Kunden- und Produkttrends aufzeigen, Möglichkeiten zur Optimierung der Infrastruktur finden oder umsetzbare Erkenntnisse für eine Vielzahl von Anwendungsfällen liefern.

Über Splunk SmartStore

Splunk SmartStore erweitert die Vorteile der Splunk-Architektur und vereinfacht gleichzeitig die kostengünstige Skalierung. Durch die Entkopplung von Rechen- und Speicherressourcen entstehen für E/A optimierte Indexerknoten mit deutlich reduziertem Speicherbedarf, da sie nur eine Teilmenge der Daten als Cache speichern. Sie müssen keine zusätzlichen Rechen- oder Speicherressourcen hinzufügen, wenn nur eine dieser Ressourcen erforderlich ist, wodurch Sie erhebliche Kosteneinsparungen erzielen können. Sie können kostengünstigen und leicht skalierbaren S3-basierten Objektspeicher verwenden, der die Umgebung weiter vereinfacht, die Kosten senkt und Ihnen die Verwaltung eines größeren Datensatzes ermöglicht.

Splunk SmartStore bietet Unternehmen einen erheblichen Mehrwert, unter anderem durch:

- Senkung der Speicherkosten durch Verschieben von warmen Daten in den kostenoptimierten S3-Objektspeicher
- Nahtlose Skalierung durch Entkopplung von Speicher und Rechenleistung
- Vereinfachung der Geschäftskontinuität durch Nutzung robuster Cloud-nativer Speicher

Lösungsübersicht

Auf dieser Seite werden die Komponenten beschrieben, die zur Vervollständigung dieser Lösung verwendet werden, darunter NetApp StorageGRID, Splunk Enterprise und Splunk SmartStore.

NetApp StorageGRID

NetApp StorageGRID ist eine leistungsstarke und kostengünstige Objektspeicherplattform. Es bietet intelligentes, richtliniengesteuertes globales Datenmanagement mithilfe einer verteilten, knotenbasierten Grid-Architektur. Es vereinfacht die Verwaltung von Petabytes unstrukturierter Daten und Milliarden von Objekten durch seinen allgegenwärtigen globalen Objekt-Namespace in Kombination mit ausgefeilten Datenverwaltungsfunktionen. Der Objektzugriff mit einem einzigen Aufruf erstreckt sich über mehrere Standorte und vereinfacht Hochverfügbarkeitsarchitekturen, während gleichzeitig ein kontinuierlicher Objektzugriff unabhängig von Standort- oder Infrastrukturausfällen gewährleistet wird.

Durch Multitenancy können mehrere Cloud- und Unternehmensanwendungen für unstrukturierte Daten sicher innerhalb desselben Grids verwaltet werden, wodurch sich der ROI und die Anwendungsfälle für StorageGRID

erhöhen. Mithilfe von metadatengesteuerten Objektlebenszyklusrichtlinien können mehrere Service-Level erstellt werden, wodurch Haltbarkeit, Schutz, Leistung und Lokalität über mehrere geografische Regionen hinweg optimiert werden. Benutzer können Richtlinien anpassen und die Datenlandschaft unterbrechungsfrei neu ausrichten, wenn sich ihre Anforderungen ändern.

SmartStore nutzt StorageGRID als Remote-Speicherebene und ermöglicht Kunden die Bereitstellung mehrerer geografisch verteilter Sites für robuste Verfügbarkeit und Haltbarkeit, dargestellt als einzelner Objekt-Namespace. Dadurch kann Splunk SmartStore die hohe Leistung und die hohe Kapazität von StorageGRID nutzen und die Möglichkeit nutzen, mithilfe einer einzigen URL auf Hunderte von Knoten an mehreren physischen Standorten zu skalieren, um mit den Objekten zu interagieren. Diese einzelne URL ermöglicht außerdem unterbrechungsfreie Speichererweiterungen, Upgrades und Reparaturen, auch über einen einzelnen Standort hinaus. Die einzigartige Datenverwaltungsrichtlinien-Engine von StorageGRID bietet optimierte Leistungs- und Haltbarkeitsniveaus sowie die Einhaltung der Anforderungen an die Datenlokalität.

Splunk Enterprise

Splunk, ein führendes Unternehmen im Bereich der Erfassung und Analyse maschinengenerierter Daten, trägt durch seine operativen Analysefunktionen zur Vereinfachung und Modernisierung der IT bei. Darüber hinaus wird es auf Anwendungsfälle in den Bereichen Geschäftsanalyse, Sicherheit und IoT ausgeweitet. Speicher ist ein entscheidender Faktor für die erfolgreiche Bereitstellung der Splunk-Software.

Maschinengenerierte Daten sind die am schnellsten wachsende Art von Big Data. Das Format ist unvorhersehbar und stammt aus vielen verschiedenen Quellen, oft mit hoher Geschwindigkeit und in großen Mengen. Diese Arbeitslastmerkmale werden oft als digitale Abgase bezeichnet. Splunk SmartStore hilft dabei, diese Daten zu verstehen und bietet intelligentes Daten-Tiering für die optimierte Platzierung heißer und warmer Daten auf der kostengünstigsten Speicherebene.

Splunk SmartStore

Splunk SmartStore ist eine Indexerfunktion, die Objektspeicher (auch als Remote-Speicher oder Remote-Speicherebenen bezeichnet) wie StorageGRID verwendet, um warme Daten mithilfe des S3-Protokolls zu speichern.

Wenn das Datenvolumen einer Bereitstellung zunimmt, übersteigt der Bedarf an Speicher in der Regel den Bedarf an Computerressourcen. Mit SmartStore können Sie Ihren Indexerspeicher und Ihre Rechenressourcen kostengünstig verwalten, indem Sie Rechenleistung und Speicher separat skalieren.

SmartStore führt eine Remote-Speicherebene unter Verwendung des S3-Protokolls und eines Cache-Managers ein. Diese Funktionen ermöglichen die Speicherung von Daten entweder lokal auf Indexern oder im Remote-Speicher. Der Cache-Manager, der sich auf dem Indexer befindet, verwaltet die Datenbewegung zwischen dem Indexer und der Remote-Speicherebene. Daten werden zusammen mit Bucket-Metadaten in Buckets (Hot und Warm) gespeichert.

Mit SmartStore können Sie den Speicherbedarf des Indexers auf ein Minimum reduzieren und E/A-optimierte Rechenressourcen auswählen, da sich die meisten Daten auf der Remote-Speicherebene befinden. Der Indexer verwaltet einen lokalen Cache, der die minimale Datenmenge darstellt, die zum Zurückgeben der angeforderten und vorhergesagten Ergebnisse erforderlich ist. Der lokale Cache enthält Hot Buckets, Kopien von Warm Buckets, die an aktiven oder kürzlich durchgeführten Suchvorgängen beteiligt sind, und Bucket-Metadaten.

Splunk SmartStore mit StorageGRID ermöglicht es Kunden, die Umgebung mit leistungsstarkem und kostengünstigem Remote-Speicher schrittweise zu skalieren und gleichzeitig der Gesamtlösung ein hohes Maß an Elastizität zu verleihen. Auf diese Weise können Kunden jederzeit beliebige Komponenten (Hot Storage und/oder Warm S3 Storage) in beliebiger Menge hinzufügen, unabhängig davon, ob sie mehr Indexer benötigen, die Datenaufbewahrung ändern oder die Aufnahmerate ohne Unterbrechung erhöhen möchten.

Vorteile dieser Lösung

Die Lösung ermöglicht das Hinzufügen von Rechen-, Hot-Storage- oder S3-Ressourcen, um die wachsende Nachfrage hinsichtlich der Anzahl der Benutzer oder der Aufnahmerate bei Bereitstellungen an einem oder mehreren Standorten zu erfüllen.

- **Leistung.** Die Kombination aus Splunk SmartStore und NetApp StorageGRID ermöglicht eine schnelle Migration von Daten zwischen Hot Buckets und Warm Buckets mithilfe von Objektspeicher. StorageGRID beschleunigt den Migrationsprozess, indem es eine schnelle Leistung für große Objekt-Workloads bietet.
- **Multisite-fähig.** Die verteilte Architektur von StorageGRID ermöglicht Splunk SmartStore die Ausweitung von Bereitstellungen auf einzelne und mehrere Standorte über einen einzigen globalen Namespace, in dem von jedem Standort aus auf die Daten zugegriffen werden kann, unabhängig davon, wo sich die Daten befinden.
- **Verbesserte Skalierbarkeit.** Skalieren Sie Speicherressourcen unabhängig von Rechenressourcen, um den sich entwickelnden Bedürfnissen und Anforderungen in Ihrer Splunk-Umgebung gerecht zu werden und so für verbesserte Gesamtbetriebskosten zu sorgen.
- **Kapazität.** Bewältigen Sie schnell wachsende Volumina bei der Splunk-Bereitstellung mit StorageGRID , indem Sie einen einzelnen Namespace auf über 560 PB skalieren.
- **Datenverfügbarkeit.** Optimieren Sie Datenverfügbarkeit, Leistung, geografische Verteilung, Aufbewahrung, Schutz und Speicherkosten mit metadatengesteuerten Richtlinien, die sich dynamisch an die Entwicklung des Geschäftswerts Ihrer Daten anpassen können.

Steigern Sie die Leistung mit dem SmartStore-Cache, einer Komponente des Indexers, die die Übertragung von Bucket-Kopien zwischen lokalem (Hot) und Remote-Speicher (Warm) übernimmt. Die Splunk-Dimensionierung für diese Lösung basiert auf der "Richtlinien von Splunk" . Die Lösung ermöglicht das Hinzufügen von Rechen-, Hot-Storage- oder S3-Ressourcen, um die wachsende Nachfrage hinsichtlich der Anzahl der Benutzer oder der Aufnahmerate bei Bereitstellungen an einem oder mehreren Standorten zu erfüllen.

Splunk-Architektur

In diesem Abschnitt wird die Splunk-Architektur beschrieben, einschließlich wichtiger Definitionen, verteilter Splunk-Bereitstellungen, Splunk SmartStore, Datenfluss, Hardware- und Softwareanforderungen, Anforderungen für Einzel- und Multisite-Umgebungen usw.

Wichtige Definitionen

In den nächsten beiden Tabellen sind die Splunk- und NetApp Komponenten aufgeführt, die in der verteilten Splunk-Bereitstellung verwendet werden.

Diese Tabelle listet die Splunk-Hardwarekomponenten für die verteilte Splunk Enterprise-Konfiguration auf.

Splunk-Komponente	Aufgabe
Indexer	Repository für Splunk Enterprise-Daten
Universal-Spediteur	Verantwortlich für die Aufnahme und Weiterleitung von Daten an die Indexer

Splunk-Komponente	Aufgabe
Suchkopf	Das Benutzer-Frontend, das zum Suchen von Daten in Indexern verwendet wird
Cluster-Master	Verwaltet die Splunk-Installation von Indexern und Suchköpfen
Überwachungskonsole	Zentralisiertes Überwachungstool für die gesamte Bereitstellung
Lizenzmaster	License Master kümmert sich um die Splunk Enterprise-Lizenzierung
Bereitstellungsserver	Aktualisiert Konfigurationen und verteilt Apps an die Verarbeitungskomponente
Speicherkomponente	Aufgabe
NetApp AFF	Zur Verwaltung von Hot-Tier-Daten wird ein vollständiger Flash-Speicher verwendet. Auch als lokaler Speicher bekannt.
NetApp StorageGRID	S3-Objektspeicher zur Verwaltung von Warm-Tier-Daten. Wird von SmartStore verwendet, um Daten zwischen der heißen und warmen Ebene zu verschieben. Auch als Remote-Speicher bekannt.

Diese Tabelle listet die Komponenten der Splunk-Speicherarchitektur auf.

Splunk-Komponente	Aufgabe	Verantwortliche Komponente
SmartStore	Bietet Indexern die Möglichkeit, Daten vom lokalen Speicher in den Objektspeicher zu verschieben.	Splunk
Heiß	Der Landeplatz, an dem Universal Forwarder neu geschriebene Daten platzieren. Der Speicher ist beschreibbar und die Daten sind durchsuchbar. Diese Datenebene besteht normalerweise aus SSDs oder schnellen HDDs.	ONTAP
Cache-Manager	Verwaltet den lokalen Cache der indizierten Daten, ruft bei einer Suche warme Daten aus dem Remote-Speicher ab und entfernt am wenigsten häufig verwendete Daten aus dem Cache.	SmartStore

Splunk-Komponente	Aufgabe	Verantwortliche Komponente
Warm	Die Daten werden logisch in den Bucket gerollt und zunächst vom Hot-Tier in den Warm-Tier umbenannt. Die Daten innerhalb dieser Ebene sind geschützt und können, wie die Hot-Tier-Ebene, aus SSDs oder HDDs mit größerer Kapazität bestehen. Sowohl inkrementelle als auch vollständige Backups werden mithilfe gängiger Datenschutzlösungen unterstützt.	StorageGRID

Verteilte Splunk-Bereitstellungen

Um größere Umgebungen zu unterstützen, in denen die Daten von vielen Maschinen stammen, müssen Sie große Datenmengen verarbeiten. Wenn viele Benutzer die Daten durchsuchen müssen, können Sie die Bereitstellung skalieren, indem Sie Splunk Enterprise-Instanzen auf mehrere Maschinen verteilen. Dies wird als verteilte Bereitstellung bezeichnet.

In einer typischen verteilten Bereitstellung führt jede Splunk Enterprise-Instanz eine spezielle Aufgabe aus und befindet sich auf einer von drei Verarbeitungsebenen, die den Hauptverarbeitungsfunktionen entsprechen.

In der folgenden Tabelle sind die Verarbeitungsebenen von Splunk Enterprise aufgeführt.

Stufe	Komponente	Beschreibung
Dateneingabe	Spediteur	Ein Forwarder verbraucht Daten und leitet die Daten dann an eine Gruppe von Indexern weiter.
Indizierung	Indexer	Ein Indexer indiziert eingehende Daten, die er normalerweise von einer Gruppe von Weiterleitungen erhält. Der Indexer wandelt die Daten in Ereignisse um und speichert die Ereignisse in einem Index. Der Indexer durchsucht die indexierten Daten auch als Antwort auf Suchanfragen eines Suchkopfs.
Suchverwaltung	Suchkopf	Ein Suchkopf dient als zentrale Ressource für die Suche. Die Suchköpfe in einem Cluster sind austauschbar und haben von jedem Mitglied des Suchkopfclusters aus Zugriff auf dieselben Suchvorgänge, Dashboards, Wissensobjekte usw.

In der folgenden Tabelle sind die wichtigen Komponenten aufgeführt, die in einer verteilten Splunk Enterprise-Umgebung verwendet werden.

Komponente	Beschreibung	Verantwortung
Index-Cluster-Master	Koordiniert Aktivitäten und Updates eines Indexer-Clusters	Indexverwaltung
Indexcluster	Gruppe von Splunk Enterprise-Indexern, die so konfiguriert sind, dass sie Daten untereinander replizieren	Indizierung
Suchkopf-Deployer	Verarbeitet die Bereitstellung und Aktualisierung des Cluster-Masters	Suchkopfverwaltung
Suchkopfcluster	Gruppe von Suchköpfen, die als zentrale Ressource für die Suche dient	Suchverwaltung
Lastenausgleich	Wird von Clusterkomponenten verwendet, um die steigende Nachfrage von Suchköpfen, Indexern und S3-Zielen zu bewältigen und die Last auf die Clusterkomponenten zu verteilen.	Lastmanagement für Clusterkomponenten

Entdecken Sie die folgenden Vorteile der verteilten Bereitstellungen von Splunk Enterprise:

- Zugriff auf vielfältige oder verteilte Datenquellen
- Bereitstellung von Funktionen zur Bewältigung der Datenanforderungen von Unternehmen jeder Größe und Komplexität
- Erreichen Sie hohe Verfügbarkeit und stellen Sie die Notfallwiederherstellung mit Datenreplikation und Multisite-Bereitstellung sicher

Splunk SmartStore

SmartStore ist eine Indexerfunktion, die es Remote-Objektspeichern wie Amazon S3 ermöglicht, indizierte Daten zu speichern. Wenn das Datenvolumen einer Bereitstellung zunimmt, übersteigt der Bedarf an Speicher in der Regel den Bedarf an Rechenressourcen. Mit SmartStore können Sie Ihren Indexerspeicher und Ihre Rechenressourcen kostengünstig verwalten, indem Sie diese Ressourcen separat skalieren.

SmartStore führt eine Remote-Speicherebene und einen Cache-Manager ein. Diese Funktionen ermöglichen die Speicherung von Daten entweder lokal auf Indexern oder auf der Remote-Speicherebene. Der Cache-Manager verwaltet die Datenbewegung zwischen dem Indexer und der Remote-Speicherebene, die auf dem Indexer konfiguriert ist.

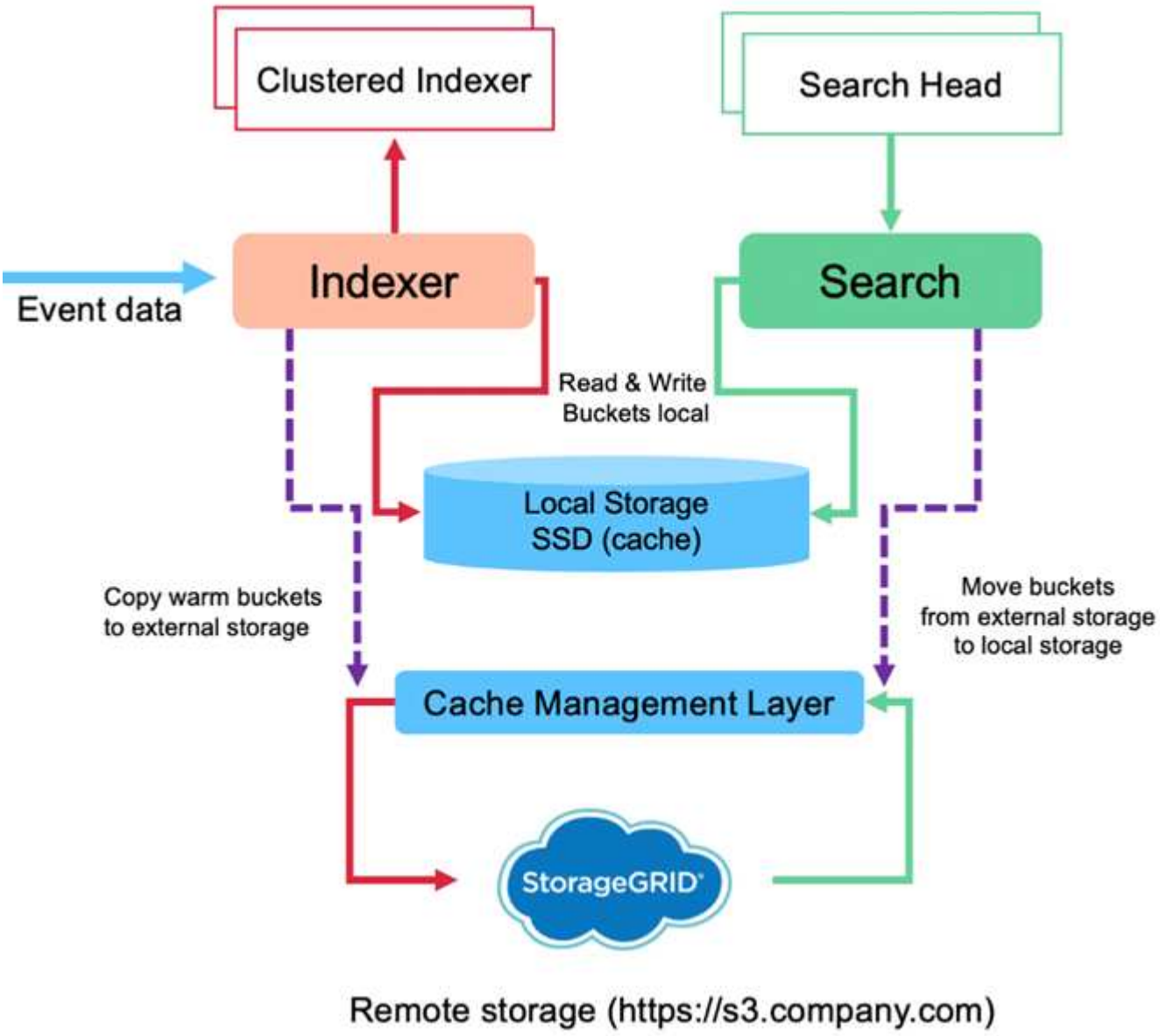
Mit SmartStore können Sie den Speicherbedarf des Indexers auf ein Minimum reduzieren und E/A-optimierte Rechenressourcen auswählen. Die meisten Daten befinden sich auf dem Remote-Speicher. Der Indexer verwaltet einen lokalen Cache, der eine minimale Datenmenge enthält: Hot Buckets, Kopien von Warm Buckets, die an aktiven oder kürzlich durchgeführten Suchvorgängen beteiligt sind, und Bucket-Metadaten.

Splunk SmartStore-Datenfluss

Wenn eingehende Daten aus verschiedenen Quellen die Indexer erreichen, werden die Daten indiziert und lokal in einem Hot Bucket gespeichert. Der Indexer repliziert außerdem die Hot-Bucket-Daten auf Zielindexer. Bisher ist der Datenfluss identisch mit dem Datenfluss für Nicht-SmartStore-Indizes.

Wenn der heiße Eimer ins Warme rollt, divergiert der Datenfluss. Der Quellindexer kopiert den Warm Bucket in den Remote-Objektspeicher (Remote-Speicherebene), während die vorhandene Kopie in seinem Cache verbleibt, da Suchvorgänge häufig über kürzlich indizierte Daten ausgeführt werden. Die Zielindexer löschen jedoch ihre Kopien, da der Remotespeicher eine hohe Verfügbarkeit bietet, ohne dass mehrere lokale Kopien verwaltet werden müssen. Die Masterkopie des Buckets befindet sich jetzt im Remote-Speicher.

Das folgende Bild zeigt den Datenfluss von Splunk SmartStore.



Der Cache-Manager auf dem Indexer ist für den SmartStore-Datenfluss von zentraler Bedeutung. Es ruft bei Bedarf Kopien von Buckets aus dem Remote-Speicher ab, um Suchanfragen zu verarbeiten. Außerdem werden ältere oder weniger häufig durchsuchte Kopien von Buckets aus dem Cache entfernt, da die Wahrscheinlichkeit, dass sie an Suchvorgängen teilnehmen, mit der Zeit abnimmt.

Die Aufgabe des Cache-Managers besteht darin, die Nutzung des verfügbaren Caches zu optimieren und gleichzeitig sicherzustellen, dass Suchvorgänge sofortigen Zugriff auf die benötigten Buckets haben.

Softwareanforderungen

In der folgenden Tabelle sind die Softwarekomponenten aufgeführt, die zur Implementierung der Lösung erforderlich sind. Die bei der Implementierung der Lösung verwendeten Softwarekomponenten können je nach Kundenanforderungen variieren.

Produktfamilie	Produktname	Produktversion	Betriebssystem
NetApp StorageGRID	StorageGRID -Objektspeicher	11,6	n/a
CentOS	CentOS	8,1	CentOS 7.x
Splunk Enterprise	Splunk Enterprise mit SmartStore	8.0.3	CentOS 7.x

Anforderungen für Einzel- und Multisite-Standorte

In einer Enterprise-Splunk-Umgebung (mittlere und große Bereitstellungen), in der die Daten von vielen Maschinen stammen und viele Benutzer die Daten durchsuchen müssen, können Sie Ihre Bereitstellung skalieren, indem Sie Splunk Enterprise-Instanzen auf einzelne und mehrere Standorte verteilen.

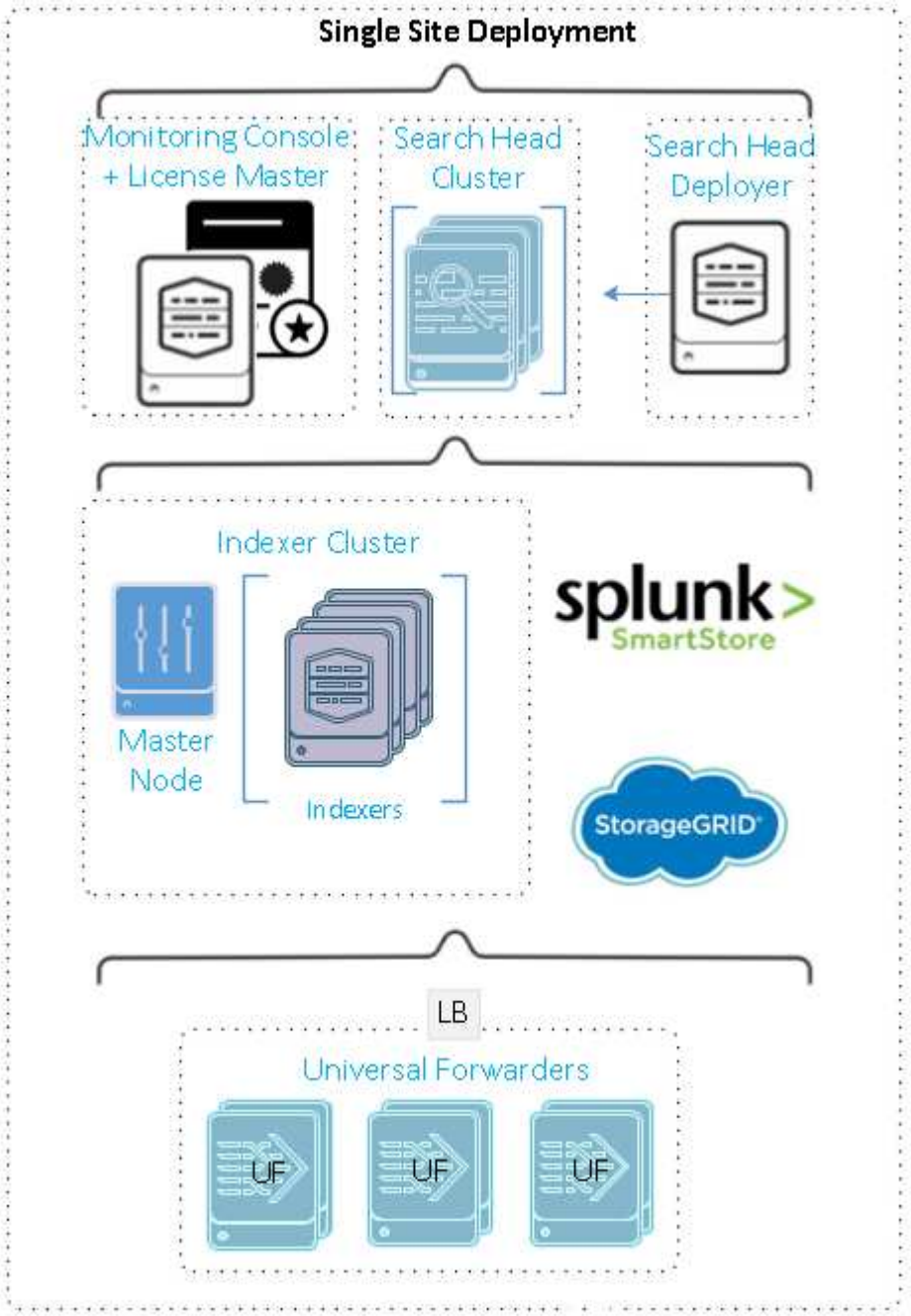
Entdecken Sie die folgenden Vorteile der verteilten Bereitstellungen von Splunk Enterprise:

- Zugriff auf vielfältige oder verteilte Datenquellen
- Bereitstellung von Funktionen zur Bewältigung der Datenanforderungen von Unternehmen jeder Größe und Komplexität
- Erreichen Sie hohe Verfügbarkeit und stellen Sie die Notfallwiederherstellung mit Datenreplikation und Multisite-Bereitstellung sicher

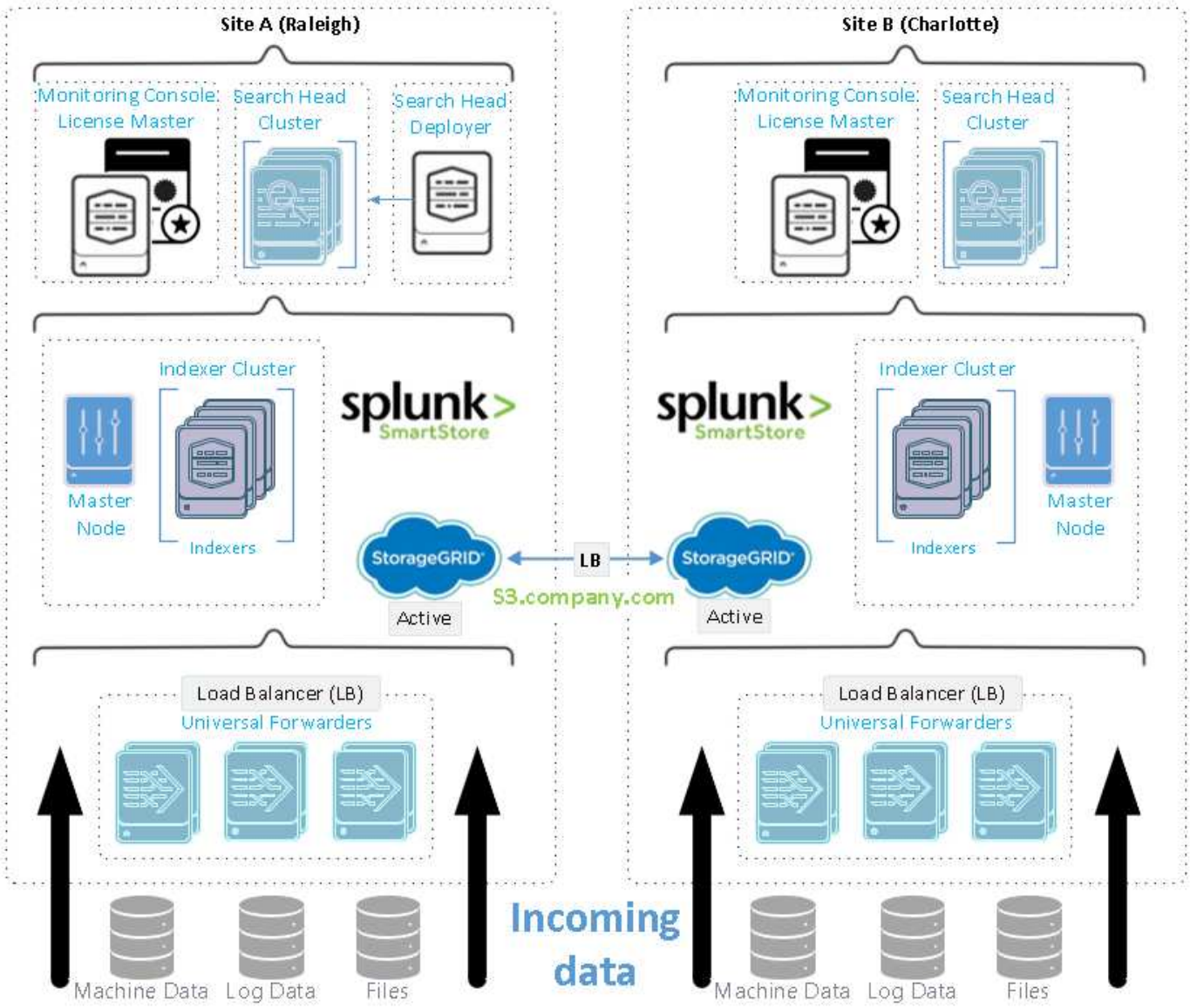
In der folgenden Tabelle sind die in einer verteilten Splunk Enterprise-Umgebung verwendeten Komponenten aufgeführt.

Komponente	Beschreibung	Verantwortung
Index-Cluster-Master	Koordiniert Aktivitäten und Updates eines Indexer-Clusters	Indexverwaltung
Indexcluster	Gruppe von Splunk Enterprise-Indexern, die so konfiguriert sind, dass sie die Daten des jeweils anderen replizieren	Indizierung
Suchkopf-Deployer	Verarbeitet die Bereitstellung und Aktualisierung des Cluster-Masters	Suchkopfverwaltung
Suchkopfcluster	Gruppe von Suchköpfen, die als zentrale Ressource für die Suche dient	Suchverwaltung
Lastenausgleichsmodule	Wird von Clusterkomponenten verwendet, um die steigende Nachfrage von Suchköpfen, Indexern und S3-Zielen zu bewältigen und die Last auf die Clusterkomponenten zu verteilen.	Lastmanagement für Clusterkomponenten

Diese Abbildung zeigt ein Beispiel für eine verteilte Bereitstellung an einem einzelnen Standort.



Diese Abbildung zeigt ein Beispiel für eine verteilte Bereitstellung an mehreren Standorten.



Hardwareanforderungen

In den folgenden Tabellen ist die Mindestanzahl an Hardwarekomponenten aufgeführt, die zur Implementierung der Lösung erforderlich sind. Die in bestimmten Implementierungen der Lösung verwendeten Hardwarekomponenten können je nach Kundenanforderungen variieren.

Unabhängig davon, ob Sie Splunk SmartStore und StorageGRID an einem oder mehreren Standorten bereitgestellt haben, werden alle Systeme über den StorageGRID GRID Manager in einer einzigen Fensteransicht verwaltet. Weitere Einzelheiten finden Sie im Abschnitt „Einfache Verwaltung mit Grid Manager“.

In dieser Tabelle ist die für einen einzelnen Standort verwendete Hardware aufgeführt.

Hardware	Menge	Scheibe	Nutzbare Kapazität	Hinweis
StorageGRID SG1000	1	n/a	n/a	Admin-Knoten und Load Balancer

Hardware	Menge	Scheibe	Nutzbare Kapazität	Hinweis
StorageGRID SG6060	4	x48, 8 TB (NL-SAS-Festplatte)	1PB	Remote-Speicher

In dieser Tabelle ist die für eine Multisite-Konfiguration verwendete Hardware (pro Site) aufgeführt.

Hardware	Menge	Scheibe	Nutzbare Kapazität	Hinweis
StorageGRID SG1000	2	n/a	n/a	Admin-Knoten und Load Balancer
StorageGRID SG6060	4	x48, 8 TB (NL-SAS-Festplatte)	1PB	Remote-Speicher

NetApp StorageGRID Load Balancer: SG1000

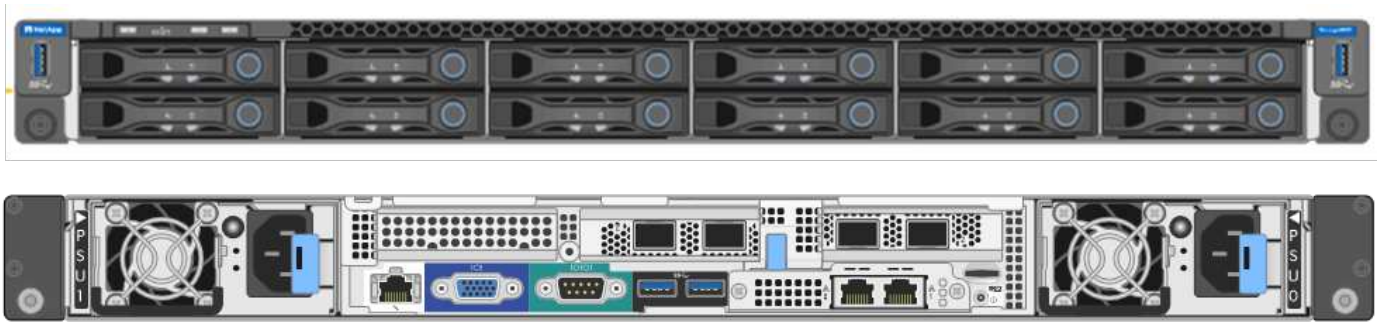
Für die Objektspeicherung ist die Verwendung eines Lastenausgleichs erforderlich, um den Cloud-Speicher-Namespace darzustellen. StorageGRID unterstützt Load Balancer von Drittanbietern führender Anbieter wie F5 und Citrix, viele Kunden entscheiden sich jedoch aufgrund seiner Einfachheit, Ausfallsicherheit und hohen Leistung für den StorageGRID -Balancer der Enterprise-Klasse. Der StorageGRID Load Balancer ist als VM, Container oder speziell entwickeltes Gerät verfügbar.

Das StorageGRID SG1000 ermöglicht die Verwendung von Hochverfügbarkeitsgruppen (HA) und intelligentem Lastenausgleich für S3-Datenpfadverbindungen. Kein anderes lokales Objektspeichersystem bietet einen angepassten Lastenausgleich.

Das SG1000-Gerät bietet die folgenden Funktionen:

- Ein Load Balancer und optional Admin-Node-Funktionen für ein StorageGRID System
- Der StorageGRID Appliance Installer vereinfacht die Bereitstellung und Konfiguration von Knoten
- Vereinfachte Konfiguration von S3-Endpunkten und SSL
- Dedizierte Bandbreite (im Gegensatz zur gemeinsamen Nutzung eines Load Balancers eines Drittanbieters mit anderen Anwendungen)
- Bis zu 4 x 100 Gbit/s aggregierte Ethernet-Bandbreite

Das folgende Bild zeigt das SG1000 Gateway Services-Gerät.



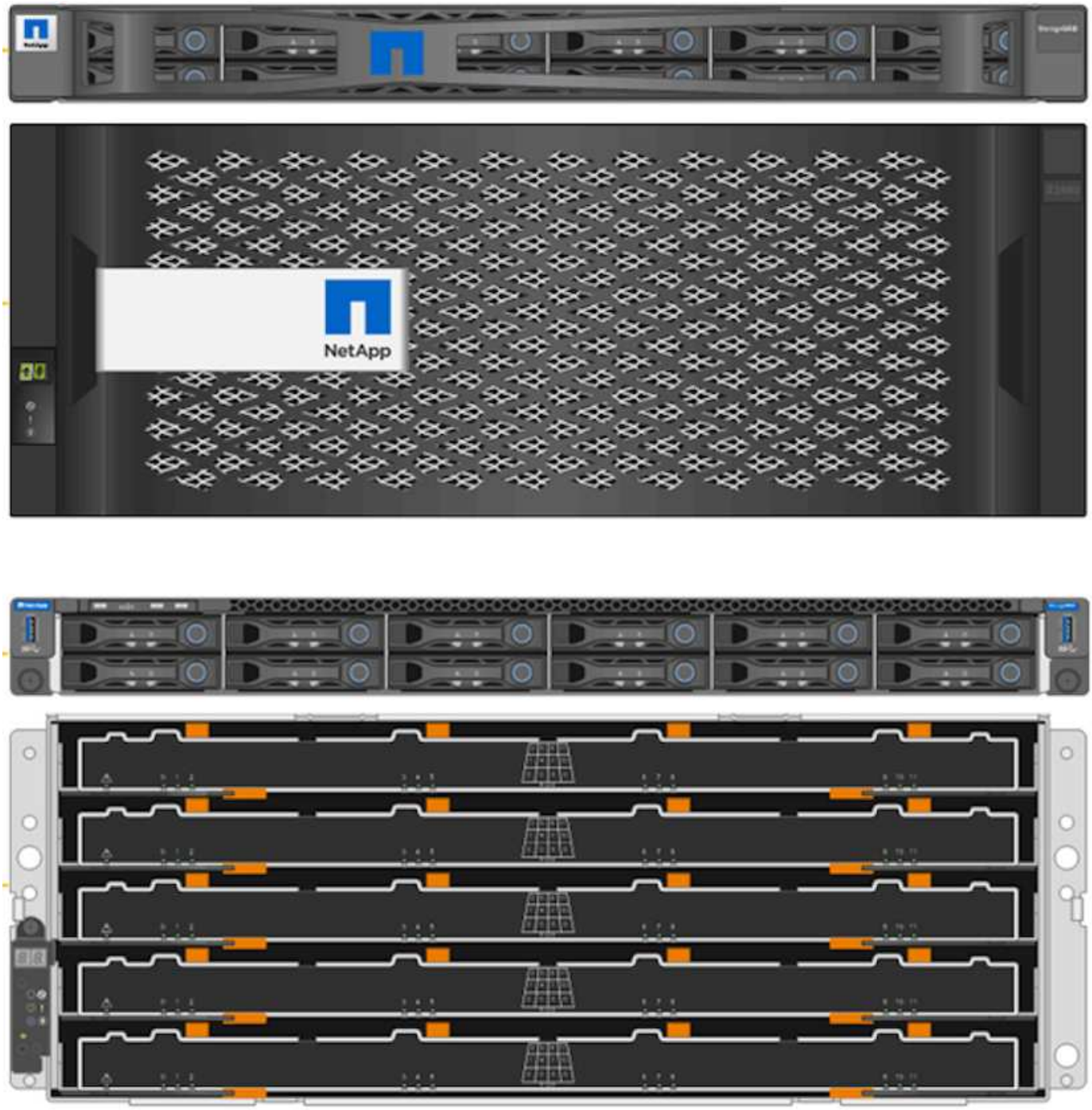
SG6060

Das StorageGRID SG6060-Gerät umfasst einen Compute Controller (SG6060) und ein Storage Controller Shelf (E-Series E2860), das zwei Storage Controller und 60 Laufwerke enthält. Dieses Gerät bietet die

folgenden Funktionen:

- Skalieren Sie bis zu 400 PB in einem einzigen Namespace.
- Bis zu 4 x 25 Gbit/s aggregierte Ethernet-Bandbreite.
- Enthält den StorageGRID Appliance Installer zur Vereinfachung der Knotenbereitstellung und -konfiguration.
- Jedes SG6060-Gerät kann über ein oder zwei zusätzliche Erweiterungsfächer für insgesamt 180 Laufwerke verfügen.
- Zwei E-Series E2800-Controller (Duplex-Konfiguration) zur Bereitstellung von Speichercontroller-Failover-Unterstützung.
- Laufwerksfach mit fünf Schubladen für sechzig 3,5-Zoll-Laufwerke (zwei Solid-State-Laufwerke und 58 NL-SAS-Laufwerke).

Das folgende Bild zeigt das Gerät SG6060.



Splunk-Design

Die folgende Tabelle listet die Splunk-Konfiguration für eine einzelne Site auf.

Splunk-Komponente	Aufgabe	Menge	Kerne	Erinnerung	Betriebssystem
Universal-Spediteur	Verantwortlich für die Aufnahme und Weiterleitung von Daten an die Indexer	4	16 Kerne	32 GB RAM	CentOS 8.1
Indexer	Verwaltet die Benutzerdaten	10	16 Kerne	32 GB RAM	CentOS 8.1
Suchkopf	Das Benutzer-Frontend durchsucht Daten in Indexern	3	16 Kerne	32 GB RAM	CentOS 8.1
Suchkopf-Deployer	Verarbeitet Updates für Suchkopfcluster	1	16 Kerne	32 GB RAM	CentOS 8.1
Cluster-Master	Verwaltet die Splunk-Installation und Indexer	1	16 Kerne	32 GB RAM	CentOS 8.1
Überwachungskonsole und Lizenzmaster	Führt eine zentrale Überwachung der gesamten Splunk-Bereitstellung durch und verwaltet Splunk-Lizenzen	1	16 Kerne	32 GB RAM	CentOS 8.1

Die folgenden Tabellen beschreiben die Splunk-Konfiguration für Multisite-Konfigurationen.

Diese Tabelle listet die Splunk-Konfiguration für eine Multisite-Konfiguration (Site A) auf.

Splunk-Komponente	Aufgabe	Menge	Kerne	Erinnerung	Betriebssystem
Universal-Spediteur	Verantwortlich für die Aufnahme und Weiterleitung der Daten an die Indexer.	4	16 Kerne	32 GB RAM	CentOS 8.1
Indexer	Verwaltet die Benutzerdaten	10	16 Kerne	32 GB RAM	CentOS 8.1

Splunk-Komponente	Aufgabe	Menge	Kerne	Erinnerung	Betriebssystem
Suchkopf	Das Benutzer-Frontend durchsucht Daten in Indexern	3	16 Kerne	32 GB RAM	CentOS 8.1
Suchkopf-Deployer	Verarbeitet Updates für Suchkopfcluster	1	16 Kerne	32 GB RAM	CentOS 8.1
Cluster-Master	Verwaltet die Splunk-Installation und Indexer	1	16 Kerne	32 GB RAM	CentOS 8.1
Überwachungskonsole und Lizenzmaster	Führt eine zentrale Überwachung der gesamten Splunk-Bereitstellung durch und verwaltet Splunk-Lizenzen.	1	16 Kerne	32 GB RAM	CentOS 8.1

Diese Tabelle listet die Splunk-Konfiguration für eine Multisite-Konfiguration (Site B) auf.

Splunk-Komponente	Aufgabe	Menge	Kerne	Erinnerung	Betriebssystem
Universal-Spediteur	Verantwortlich für die Aufnahme und Weiterleitung von Daten an die Indexer	4	16 Kerne	32 GB RAM	CentOS 8.1
Indexer	Verwaltet die Benutzerdaten	10	16 Kerne	32 GB RAM	CentOS 8.1
Suchkopf	Das Benutzer-Frontend durchsucht Daten in Indexern	3	16 Kerne	32 GB RAM	CentOS 8.1
Cluster-Master	Verwaltet die Splunk-Installation und Indexer	1	16 Kerne	32 GB RAM	CentOS 8.1

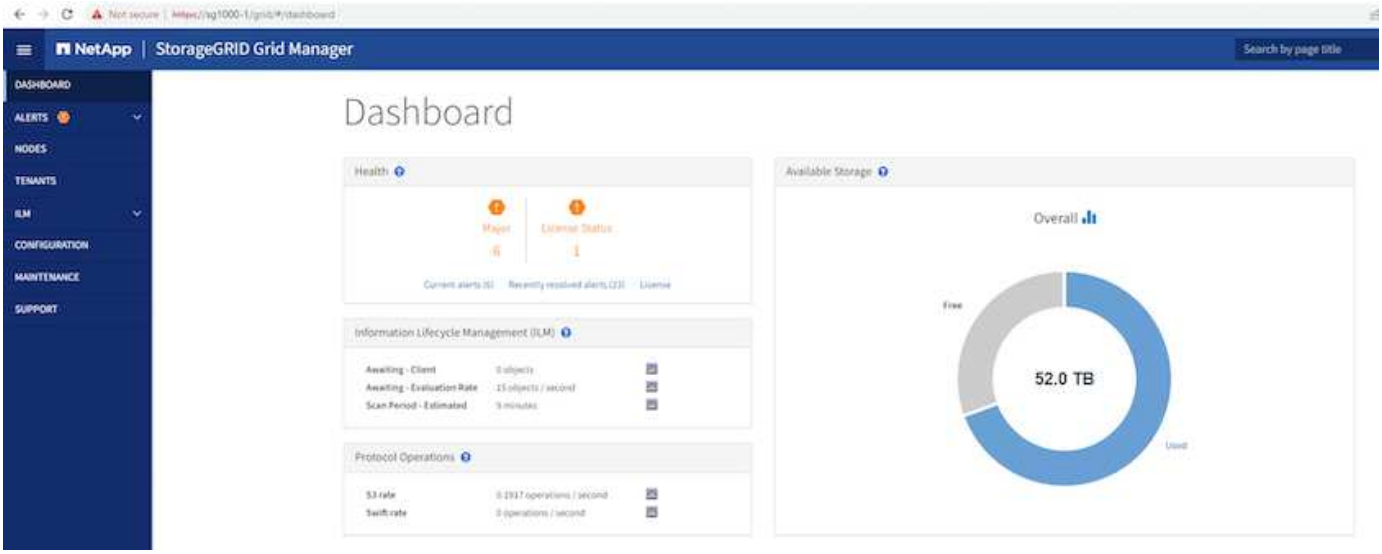
Splunk-Komponente	Aufgabe	Menge	Kerne	Erinnerung	Betriebssystem
Überwachungskonsole und Lizenzmaster	Führt eine zentrale Überwachung der gesamten Splunk-Bereitstellung durch und verwaltet Splunk-Lizenzen	1	16 Kerne	32 GB RAM	CentOS 8.1

Flexible StorageGRID -Funktionen für Splunk SmartStore

StorageGRID verfügt über eine Vielzahl von Funktionen, die Benutzer nutzen und an ihre sich ständig ändernde Umgebung anpassen können. Von der Bereitstellung bis zur Skalierung Ihres Splunk SmartStore erfordert Ihre Umgebung eine schnelle Anpassung an Änderungen und sollte Splunk nicht stören. Mit den flexiblen Datenverwaltungsrichtlinien (ILM) und Verkehrsklassifizierern (QoS) von StorageGRID können Sie Ihre Umgebung planen und anpassen.

Einfache Verwaltung mit Grid Manager

Grid Manager ist die browserbasierte grafische Benutzeroberfläche, mit der Sie Ihr StorageGRID -System an weltweit verteilten Standorten in einer einzigen Fensteransicht konfigurieren, verwalten und überwachen können, wie in der folgenden Abbildung dargestellt.



Führen Sie die folgenden Aufgaben mit der Grid Manager-Schnittstelle aus:

- Verwalten Sie global verteilte Repositories im Petabyte-Bereich mit Objekten wie Bildern, Videos und Aufzeichnungen.
- Überwachen Sie Grid-Knoten und -Dienste, um die Objektverfügbarkeit sicherzustellen.
- Verwalten Sie die Platzierung von Objektdaten im Laufe der Zeit mithilfe von Regeln für das Information

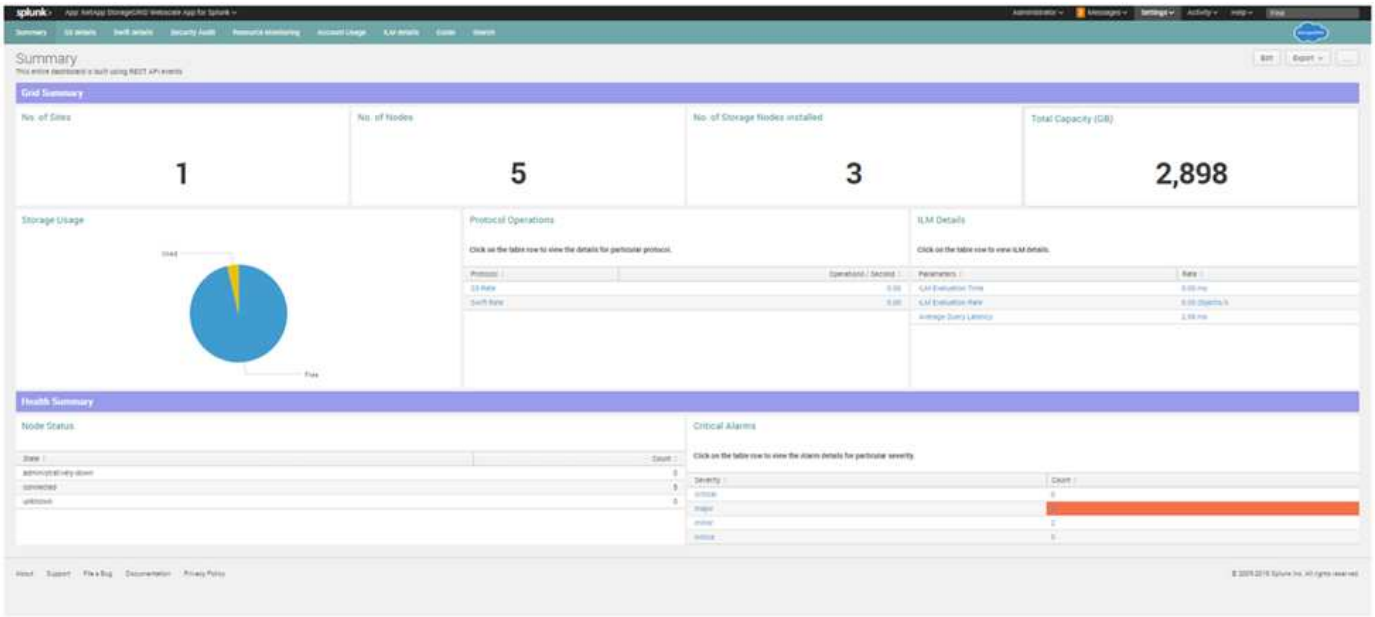
Lifecycle Management (ILM). Diese Regeln legen fest, was mit den Daten eines Objekts nach der Aufnahme geschieht, wie sie vor Verlust geschützt werden, wo und wie lange die Objektdaten gespeichert werden.

- Überwachen Sie Transaktionen, Leistung und Vorgänge innerhalb des Systems.

NetApp StorageGRID App für Splunk

Die NetApp StorageGRID App für Splunk ist eine spezielle Anwendung für Splunk Enterprise. Diese App funktioniert in Verbindung mit dem NetApp StorageGRID Add-on für Splunk. Es bietet Einblick in den StorageGRID -Zustand, Informationen zur Kontonutzung, Details zur Sicherheitsüberprüfung, Ressourcennutzung und -überwachung usw.

Das folgende Bild zeigt die StorageGRID -App für Splunk.



ILM-Richtlinien

StorageGRID verfügt über flexible Datenverwaltungsrichtlinien, die das Aufbewahren mehrerer Kopien Ihrer Objekte und die Verwendung von EC-Schemata (Erasure Coding) wie 2+1 und 4+2 (und vielen anderen) zum Speichern Ihrer Objekte je nach spezifischen Leistungs- und Datenschutzanforderungen umfassen. Da sich Arbeitslasten und Anforderungen im Laufe der Zeit ändern, ist es üblich, dass sich auch die ILM-Richtlinien im Laufe der Zeit ändern müssen. Das Ändern von ILM-Richtlinien ist eine Kernfunktion, die es StorageGRID Kunden ermöglicht, sich schnell und einfach an ihre sich ständig ändernde Umgebung anzupassen.

Performance

StorageGRID skaliert die Leistung durch Hinzufügen weiterer Knoten, bei denen es sich um VMs, Bare Metal oder speziell entwickelte Geräte wie SG5712, SG5760, SG6060 oder SGF6024 handeln kann. In unseren Tests haben wir die wichtigsten Leistungsanforderungen von SmartStore mit einem Drei-Knoten-Raster der Mindestgröße unter Verwendung des SG6060-Geräts übertroffen. Wenn Kunden ihre Splunk-Infrastruktur mit zusätzlichen Indexern skalieren, können sie weitere Speicherknoten hinzufügen, um Leistung und Kapazität zu erhöhen.

Load Balancer und Endpunktkonfiguration

Admin-Knoten in StorageGRID bieten die Grid Manager-Benutzeroberfläche (Benutzeroberfläche) und den REST-API-Endpunkt zum Anzeigen, Konfigurieren und Verwalten Ihres StorageGRID -Systems sowie Prüfprotokolle zum Verfolgen der Systemaktivität. Um einen hochverfügbaren S3-Endpunkt für den Remote-Speicher von Splunk SmartStore bereitzustellen, haben wir den StorageGRID Load Balancer implementiert, der als Dienst auf Admin-Knoten und Gateway-Knoten ausgeführt wird. Darüber hinaus verwaltet der Load Balancer auch den lokalen Datenverkehr und kommuniziert mit dem GSLB (Global Server Load Balancing), um bei der Notfallwiederherstellung zu helfen.

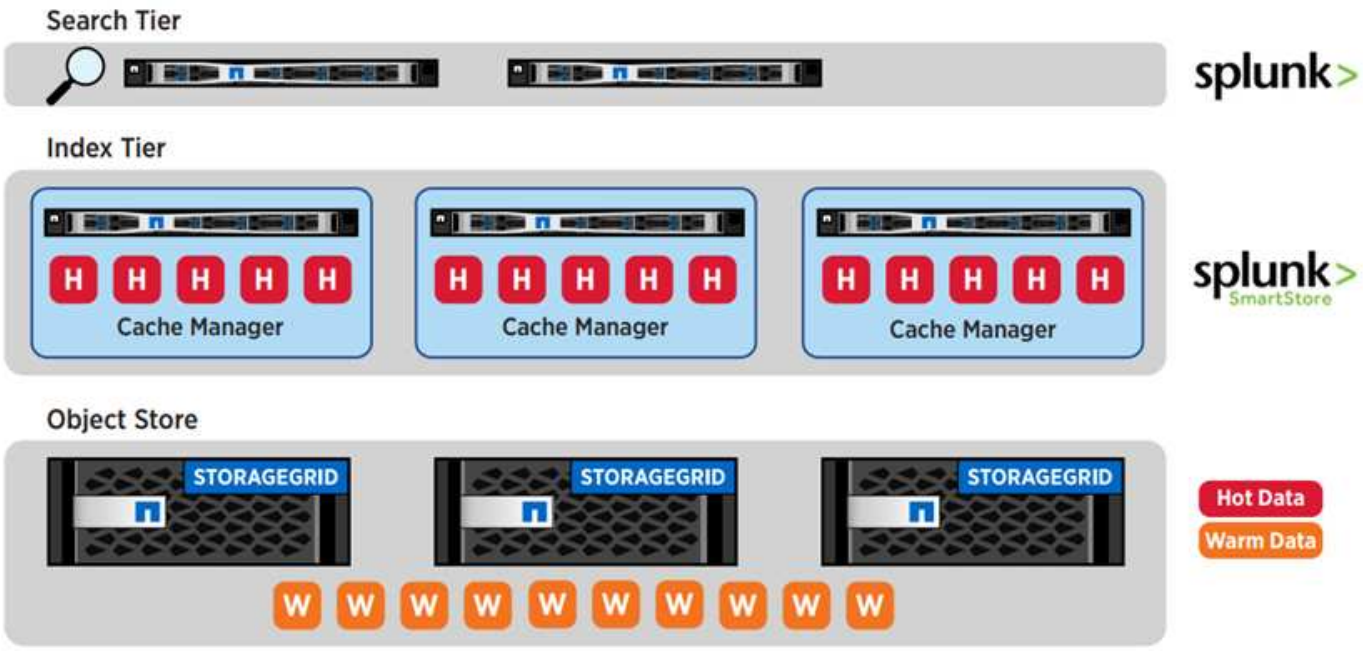
Um die Endpunktkonfiguration weiter zu verbessern, bietet StorageGRID im Admin-Knoten integrierte Richtlinien zur Verkehrsklassifizierung, ermöglicht Ihnen die Überwachung Ihres Workload-Verkehrs und die Anwendung verschiedener Quality-of-Service-Grenzwerte (QoS) auf Ihre Workloads. Richtlinien zur Verkehrsklassifizierung werden auf Endpunkte des StorageGRID Load Balancer-Dienstes für Gateway-Knoten und Admin-Knoten angewendet. Diese Richtlinien können bei der Begrenzung und Überwachung des Datenverkehrs helfen.

Intelligentes Tiering und Kosteneinsparungen

Da die Kunden die Leistungsfähigkeit und Benutzerfreundlichkeit der Splunk-Datenanalyse erkennen, möchten sie natürlich eine immer größer werdende Datenmenge indizieren. Mit der wachsenden Datenmenge wächst auch die Rechen- und Speicherinfrastruktur, die zu ihrer Verarbeitung erforderlich ist. Da auf ältere Daten weniger häufig verwiesen wird, wird es zunehmend ineffizient, die gleiche Menge an Rechenressourcen bereitzustellen und teuren Primärspeicher zu verbrauchen. Um im großen Maßstab arbeiten zu können, profitieren Kunden von der Verschiebung warmer Daten auf eine kostengünstigere Ebene, wodurch Rechenleistung und Primärspeicher für heiße Daten frei werden.

Splunk SmartStore mit StorageGRID bietet Unternehmen eine skalierbare, leistungsstarke und kostengünstige Lösung. Da SmartStore datenbewusst ist, wertet es automatisch Datenzugriffsmuster aus, um zu bestimmen, welche Daten für Echtzeitanalysen zugänglich sein müssen (Hot Data) und welche Daten in einem kostengünstigeren Langzeitspeicher verbleiben sollten (Warm Data). SmartStore verwendet die branchenübliche AWS S3-API dynamisch und intelligent und platziert Daten im von StorageGRID bereitgestellten S3-Speicher. Die flexible Scale-Out-Architektur von StorageGRID ermöglicht ein kostengünstiges Wachstum der Warm Data-Ebene nach Bedarf. Die knotenbasierte Architektur von StorageGRID stellt sicher, dass Leistungs- und Kostenanforderungen optimal erfüllt werden.

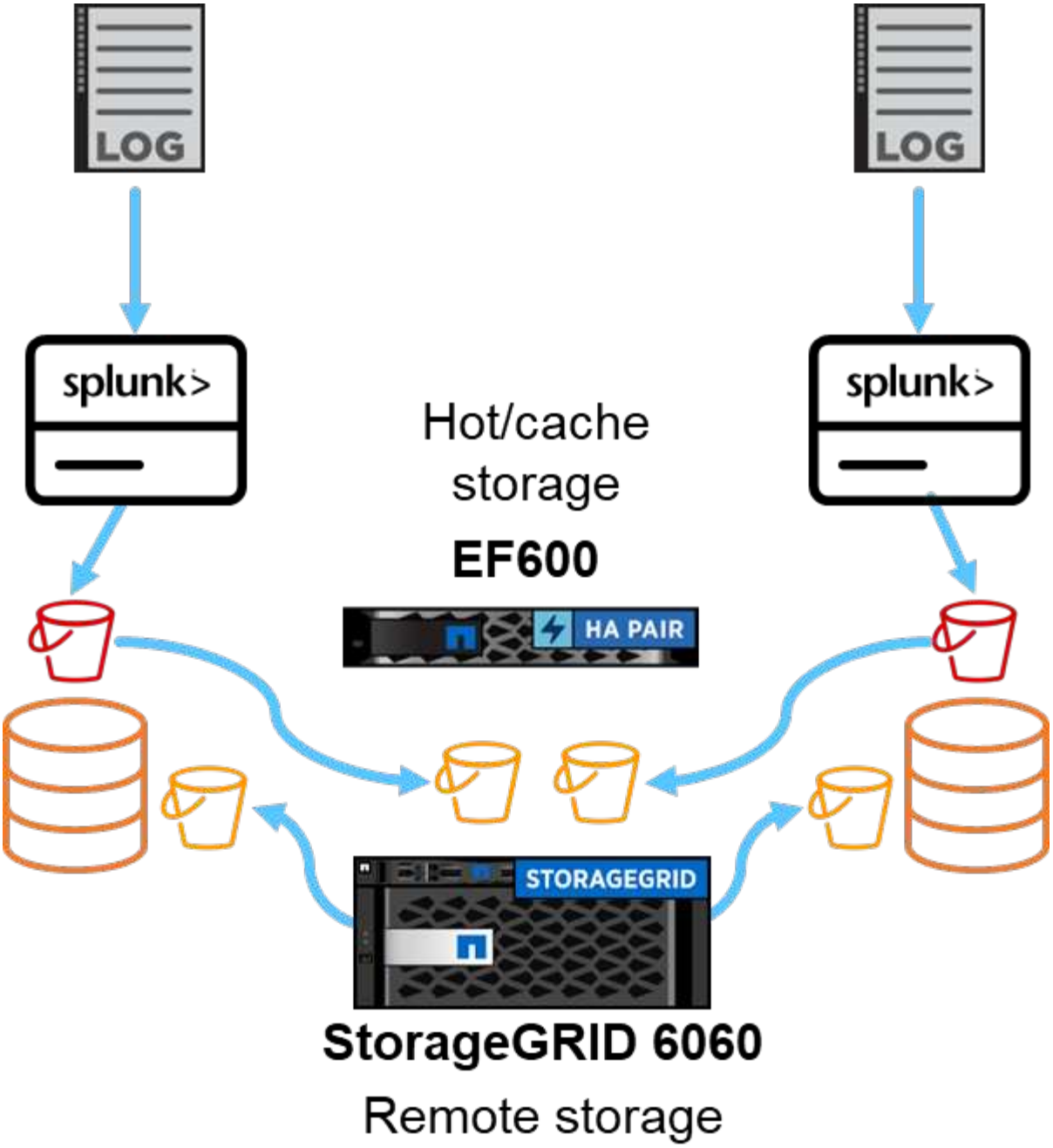
Das folgende Bild veranschaulicht die Splunk- und StorageGRID -Tiering.



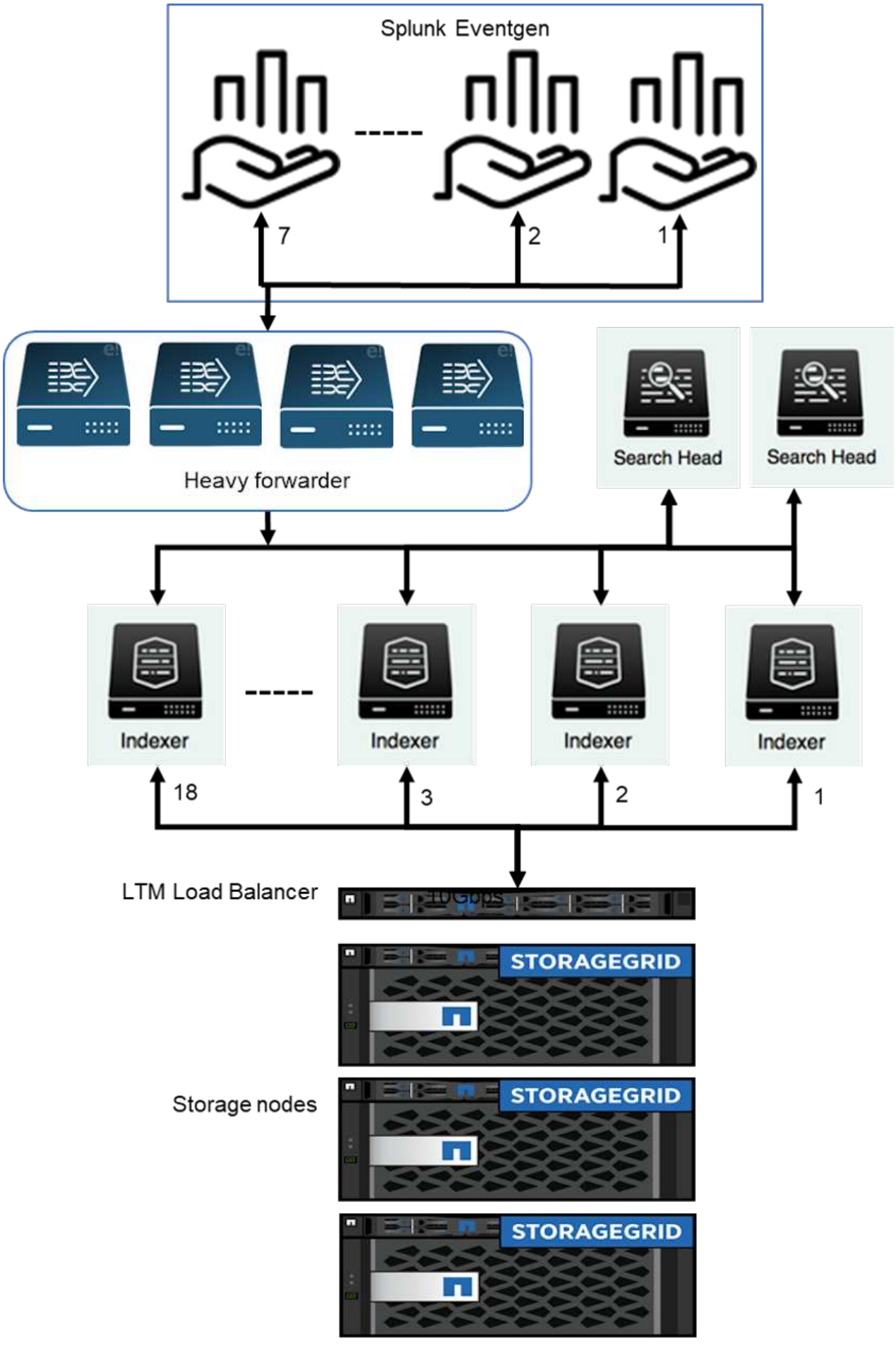
Die branchenführende Kombination von Splunk SmartStore mit NetApp StorageGRID bietet die Vorteile einer entkoppelten Architektur durch eine Full-Stack-Lösung.

SmartStore-Leistung an einem einzelnen Standort

In diesem Abschnitt wird die Leistung von Splunk SmartStore auf einem NetApp StorageGRID Controller beschrieben. Splunk SmartStore verschiebt warme Daten in den Remote-Speicher, in diesem Fall in den StorageGRID Objektspeicher bei der Leistungsvalidierung.



Wir haben EF600 für Hot-/Cache-Speicher und StorageGRID 6060 für Remote-Speicher verwendet. Für die Leistungsvalidierung haben wir die folgende Architektur verwendet. Wir haben zwei Suchköpfe, vier Heavy Forwarder zum Weiterleiten der Daten an Indexer, sieben Splunk Event Generators (Eventgens) zum Generieren der Echtzeitdaten und 18 Indexer zum Speichern der Daten verwendet.


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

Konfiguration

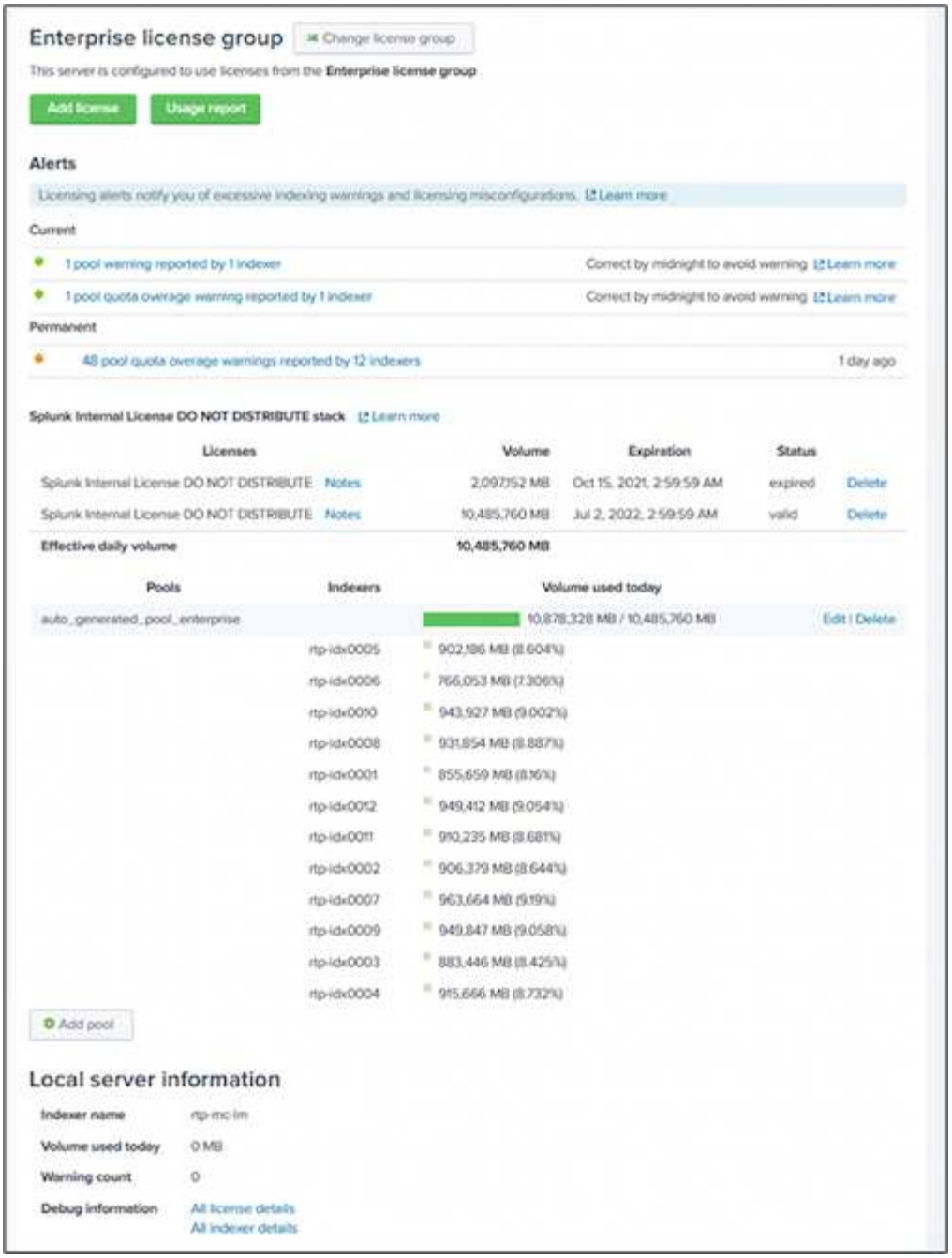
In dieser Tabelle ist die für die Leistungsvalidierung von SmartStorage verwendete Hardware aufgeführt.

Splunk-Komponente	Aufgabe	Menge	Kerne	Erinnerung	Betriebssystem
Schwerlast-Forwarder	Verantwortlich für die Aufnahme und Weiterleitung von Daten an die Indexer	4	16 Kerne	32 GB RAM	SLED 15 SP2
Indexer	Verwaltet die Benutzerdaten	18	16 Kerne	32 GB RAM	SLED 15 SP2
Suchkopf	Das Benutzer-Frontend sucht Daten in Indexern	2	16 Kerne	32 GB RAM	SLED 15 SP2
Suchkopf-Deployer	Verarbeitet Updates für Suchkopfcluster	1	16 Kerne	32 GB RAM	SLED 15 SP2
Cluster-Master	Verwaltet die Splunk-Installation und Indexer	1	16 Kerne	32 GB RAM	SLED 15 SP2
Überwachungskonsole und Lizenzmaster	Führt eine zentrale Überwachung der gesamten Splunk-Bereitstellung durch und verwaltet Splunk-Lizenzen	1	16 Kerne	32 GB RAM	SLED 15 SP2

SmartStore Remote Store-Leistungsvalidierung

Bei dieser Leistungsvalidierung haben wir den SmartStore-Cache im lokalen Speicher auf allen Indexern für 10 Tage Daten konfiguriert. Wir haben die `maxDataSize=auto` (750 MB Bucket-Größe) im Splunk-Cluster-Manager und habe die Änderungen an alle Indexer gesendet. Um die Upload-Leistung zu messen, haben wir 10 Tage lang täglich 10 TB aufgenommen und alle Hot Buckets gleichzeitig auf Warm übertragen. Außerdem haben wir den Spitzen- und Durchschnittsdurchsatz pro Instanz und Bereitstellung über das Dashboard der SmartStore Monitoring Console erfasst.

Dieses Bild zeigt die an einem Tag aufgenommenen Daten.



Wir haben den folgenden Befehl vom Cluster-Master ausgeführt (der Indexname ist `eventgen-test`). Anschließend haben wir den Spitzen- und Durchschnitts-Upload-Durchsatz pro Instanz und Bereitstellungsweite über die Dashboards der SmartStore Monitoring Console erfasst.

```
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
```

Der Cluster-Master verfügt über eine passwortlose Authentifizierung für alle Indexer (rtp-idx0001…rtp-idx0018).

Um die Download-Leistung zu messen, haben wir alle Daten aus dem Cache entfernt, indem wir die Evict-CLI mit dem folgenden Befehl zweimal ausgeführt haben.

Wir haben den folgenden Befehl vom Clustermaster aus ausgeführt und die Suche vom Suchkopf aus über 10 Tage Daten aus dem Remotespeicher von StorageGRID ausgeführt. Anschließend haben wir den Spitzen- und Durchschnitts-Upload-Durchsatz pro Instanz und Bereitstellungsweite über die Dashboards der SmartStore Monitoring Console erfasst.

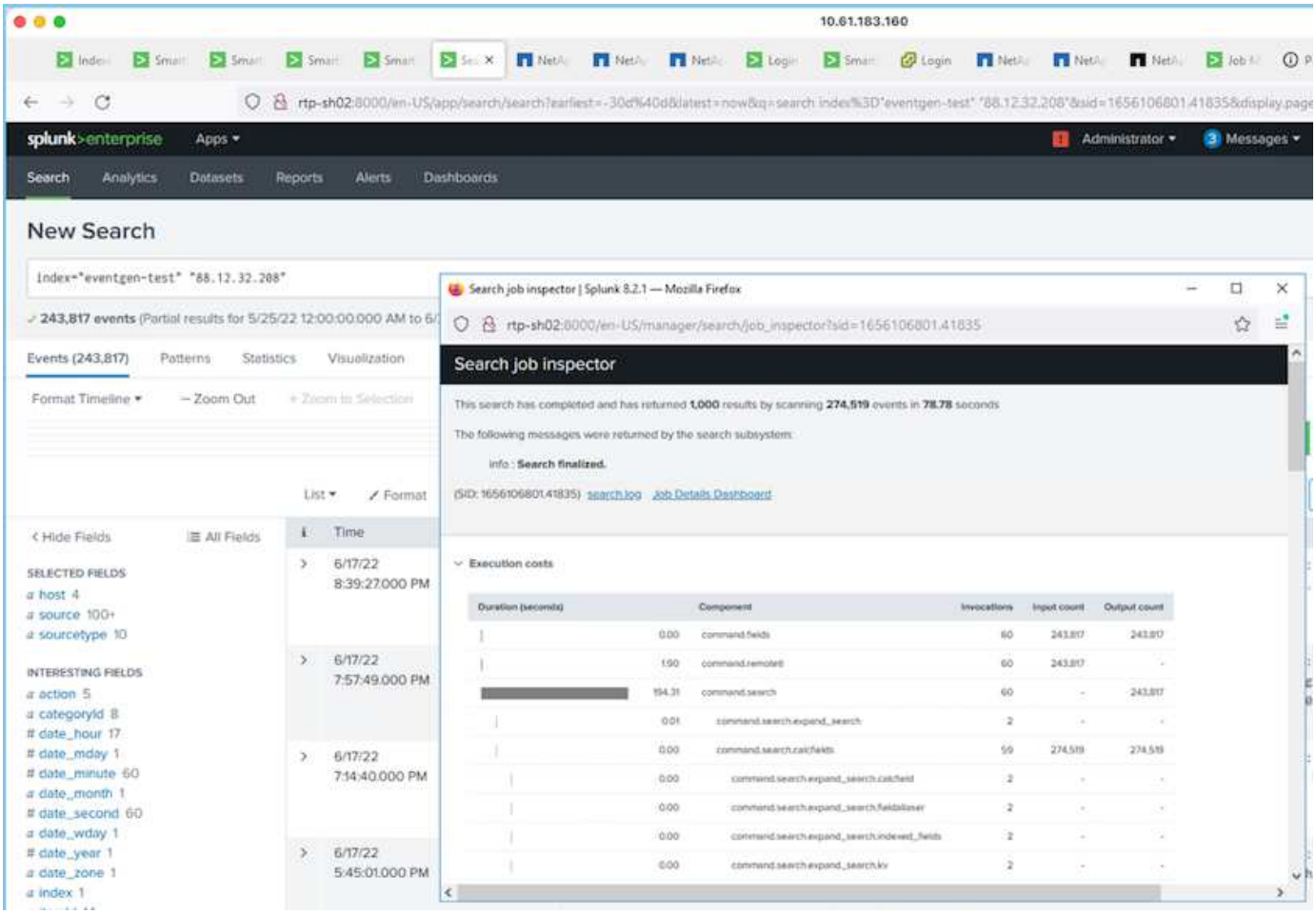
```
for i in rtp-idx0001 rtp-idx0002 rtp-idx0003 rtp-idx0004 rtp-idx0005 rtp-
idx0006 rtp-idx0007 rtp-idx0008 rtp-idx0009 rtp-idx0010 rtp-idx0011 rtp-
idx0012 rtp-idx0013 rtp-idx0014 rtp-idx0015 rtp-idx0016 rtp-idx0017 rtp-
idx0018 ; do  ssh $i " hostname;  date; /opt/splunk/bin/splunk _internal
call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

Die Indexerkonfigurationen wurden vom SmartStore-Clustermaster gepusht. Der Cluster-Master hatte die folgende Konfiguration für den Indexer.

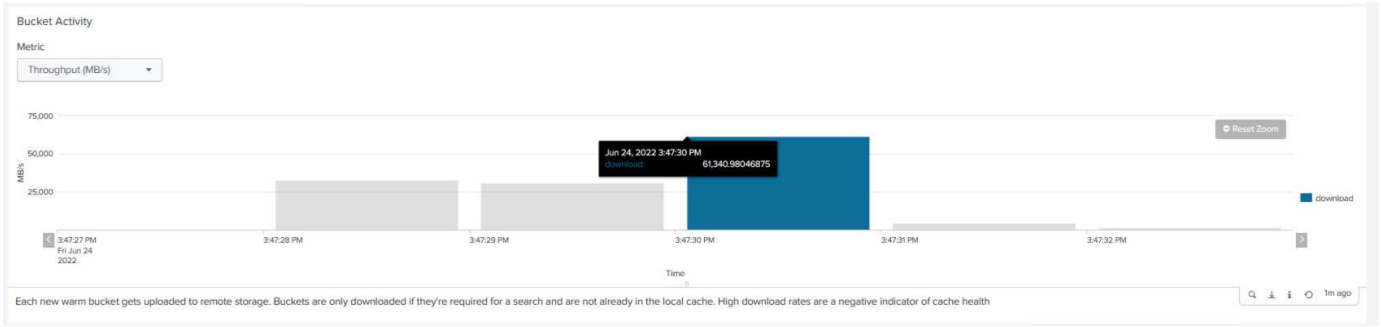
```
Rtp-cm01:~ # cat /opt/splunk/etc/master-apps/_cluster/local/indexes.conf
[default]
maxDataSize = auto
#defaultDatabase = eventgen-basic
defaultDatabase = eventgen-test
hotlist_recency_secs = 864000
repFactor = auto
[volume:remote_store]
storageType = remote
path = s3://smartstore2
remote.s3.access_key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

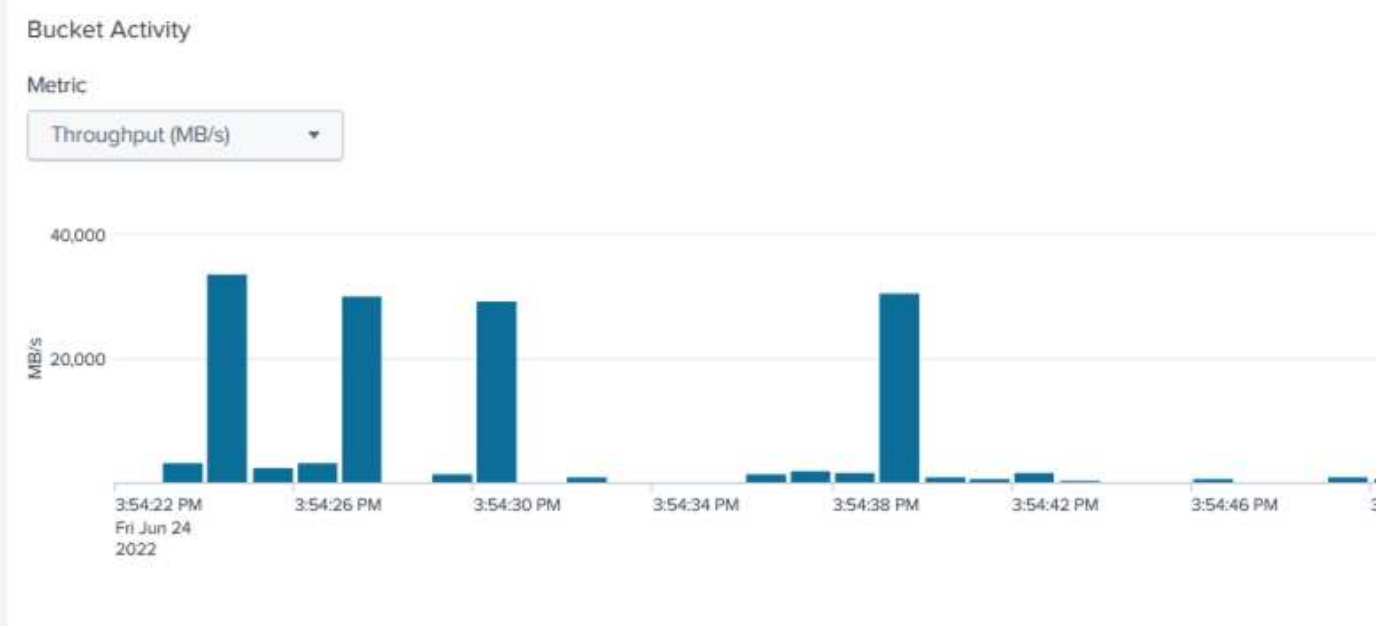
Wir haben die folgende Suchanfrage im Suchkopf ausgeführt, um die Leistungsmatrix zu erfassen.



Wir haben die Leistungsinformationen vom Cluster-Master gesammelt. Die Spitzenleistung betrug 61,34 GBps.



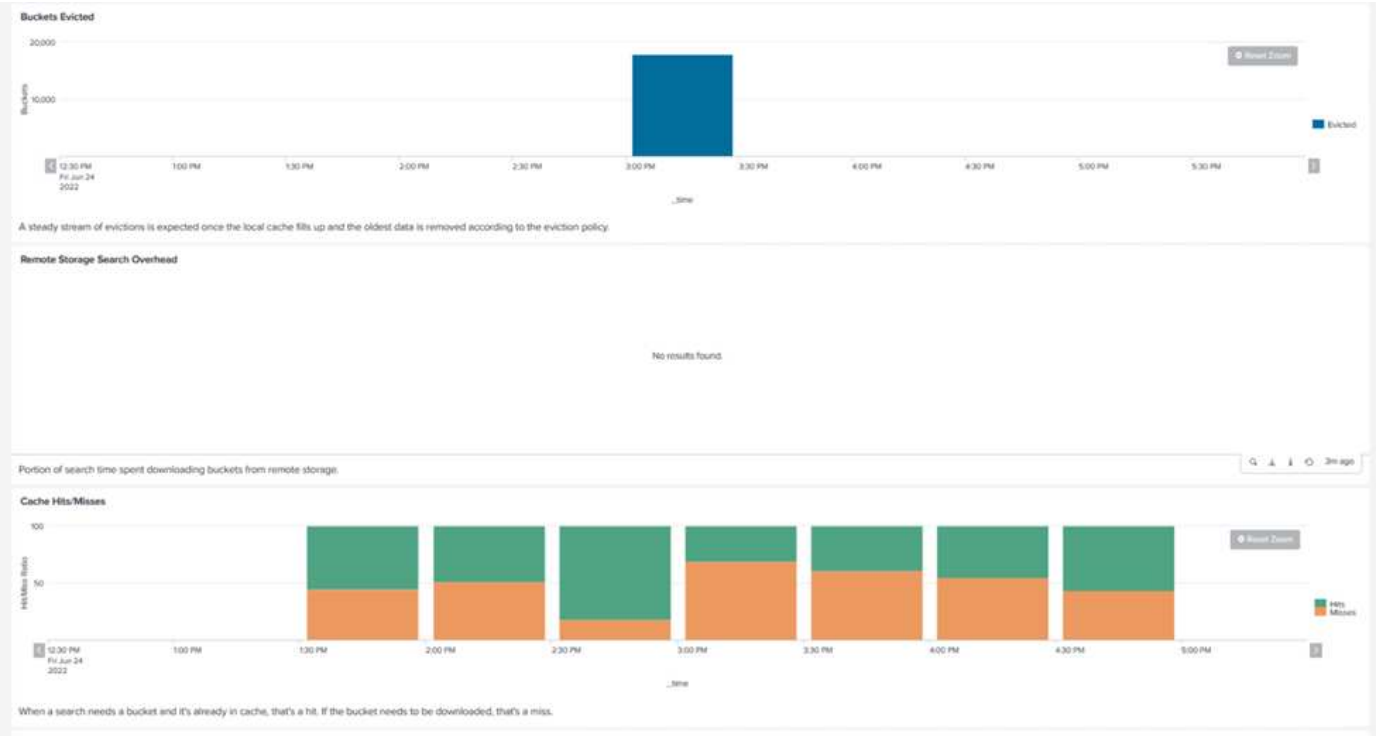
Die durchschnittliche Leistung lag bei etwa 29 GBps.



StorageGRID -Leistung

Die Leistung von SmartStore basiert auf der Suche nach bestimmten Mustern und Zeichenfolgen in großen Datenmengen. Bei dieser Validierung werden die Ereignisse generiert mit "Eventgen" auf einem bestimmten Splunk-Index (eventgen-test) über den Suchkopf, und die Anfrage geht für die meisten Abfragen an StorageGRID . Das folgende Bild zeigt die Treffer und Fehlschläge der Abfragedaten. Die Trefferdaten stammen von der lokalen Festplatte und die Fehlerdaten vom StorageGRID Controller.

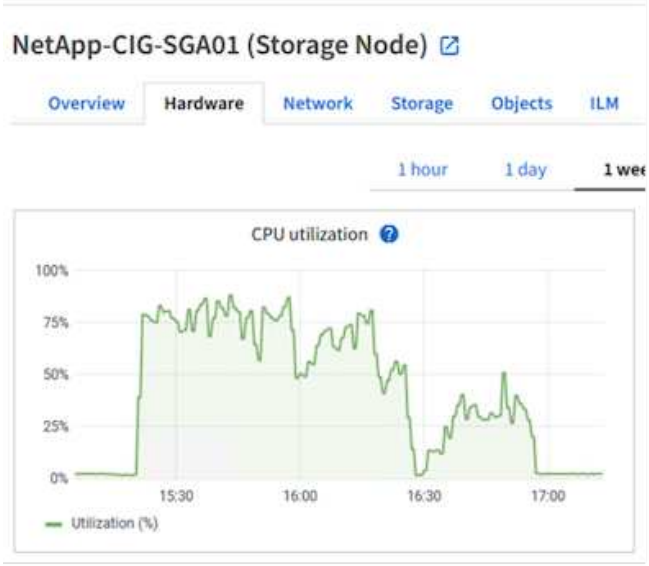
Die grüne Farbe zeigt die Trefferdaten und die orange Farbe die Fehltrefferdaten.

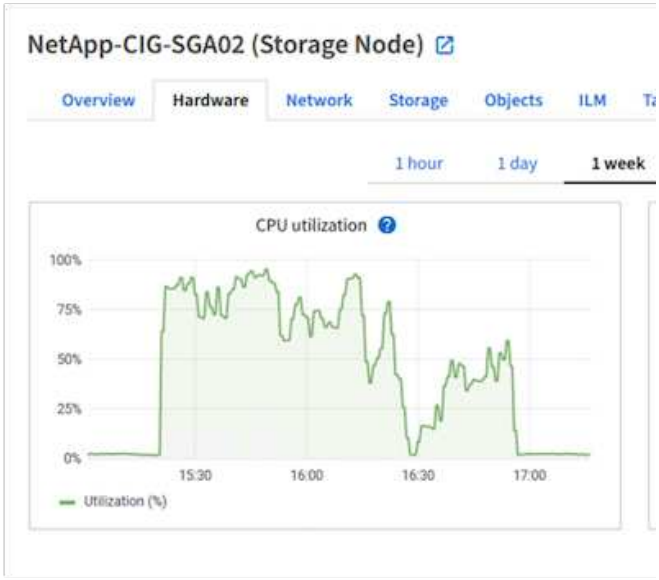


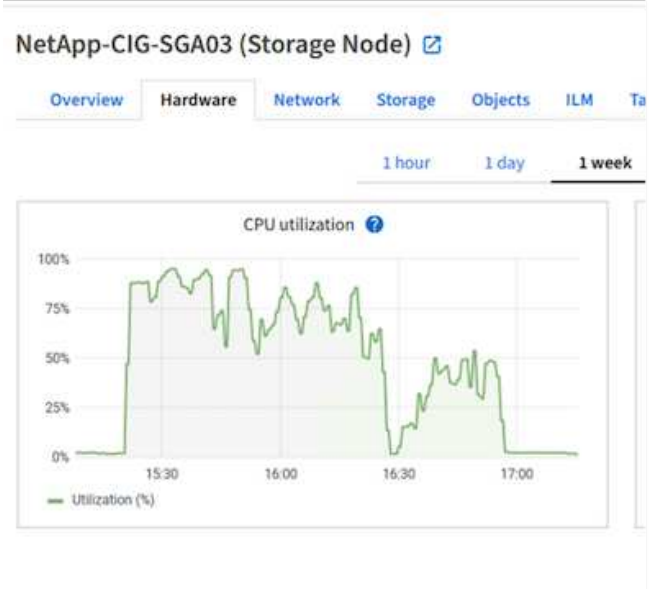
Wenn die Abfrage für die Suche auf StorageGRID ausgeführt wird, wird die Zeit für die S3-Abrufrate von StorageGRID im folgenden Bild angezeigt.

StorageGRID Hardwarenutzung

Die StorageGRID Instanz verfügt über einen Load Balancer und drei StorageGRID Controller. Die CPU-Auslastung für alle drei Controller liegt zwischen 75 % und 100 %.







SmartStore mit NetApp Storage Controller – Vorteile für den Kunden

- **Entkopplung von Rechenleistung und Speicher.** Der Splunk SmartStore entkoppelt Rechenleistung und Speicher, sodass Sie diese unabhängig voneinander skalieren können.
- **Daten auf Anfrage.** SmartStore bringt Daten bei Bedarf in die Nähe der Rechenleistung und bietet Rechen- und Speicherelastizität sowie Kosteneffizienz, um eine längere Datenaufbewahrung im großen Maßstab zu erreichen.
- **AWS S3 API-kompatibel.** SmartStore verwendet die AWS S3-API zur Kommunikation mit dem Wiederherstellungsspeicher, einem AWS S3- und S3-API-kompatiblen Objektspeicher wie StorageGRID.
- **Reduziert Speicherbedarf und Kosten.** SmartStore reduziert den Speicherbedarf für ältere Daten (warm/kalt). Es wird nur eine einzige Datenkopie benötigt, da der NetApp Speicher Datenschutz bietet und sich um Ausfälle und hohe Verfügbarkeit kümmert.
- **Hardwarefehler.** Ein Knotenausfall in einer SmartStore-Bereitstellung macht die Daten nicht unzugänglich und ermöglicht eine viel schnellere Wiederherstellung des Indexers nach einem Hardwarefehler oder Datenungleichgewicht.
- Anwendungs- und datenbewusster Cache.

- Indexer hinzufügen/entfernen und Cluster nach Bedarf einrichten/abbauen.
- Die Speicherebene ist nicht mehr an die Hardware gebunden.

Abschluss

Splunk Enterprise ist die marktführende SIEM-Lösung, die in Sicherheits-, IT- und DevOps-Teams zu besseren Ergebnissen führt. Die Nutzung von Splunk hat in den Organisationen unserer Kunden erheblich zugenommen. Daher müssen weitere Datenquellen hinzugefügt und die Daten gleichzeitig über einen längeren Zeitraum aufbewahrt werden, was die Splunk-Infrastruktur belastet.

Die Kombination aus Splunk SmartStore und NetApp StorageGRID soll Unternehmen eine skalierbare Architektur bieten, mit der sie eine verbesserte Aufnahmeleistung mit SmartStore- und StorageGRID Objektspeicher sowie eine erhöhte Skalierbarkeit für eine Splunk-Umgebung über mehrere geografische Regionen hinweg erreichen können.

Wo Sie weitere Informationen finden

Weitere Informationen zu den in diesem Dokument beschriebenen Informationen finden Sie in den folgenden Dokumenten und/oder auf den folgenden Websites:

- "NetApp StorageGRID Dokumentationsressourcen"
- "NetApp Produktdokumentation"
- "Splunk Enterprise-Dokumentation"
- "Splunk Enterprise Über SmartStore"
- "Handbuch zur verteilten Bereitstellung von Splunk Enterprise"
- "Splunk Enterprise – Verwalten von Indexern und Indexerclustern"