NetApp StorageGRID con Splunk SmartStore

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Sommario

NetApp StorageGRID con Splunk SmartStore

TR-4869: NetApp StorageGRID con Splunk SmartStore

Splunk Enterprise è la soluzione SIEM (Security Information and Event Management) leader di mercato che garantisce risultati nei team di sicurezza, IT e DevOps.

Panoramica

I volumi di dati continuano a crescere a ritmi esponenziali, creando enormi opportunità per le aziende che riescono a sfruttare questa vasta risorsa. Splunk Enterprise continua a essere adottato in una più ampia gamma di casi d'uso. Con l'aumentare dei casi d'uso, aumenta anche la quantità di dati che Splunk Enterprise acquisisce ed elabora. L'architettura tradizionale di Splunk Enterprise è un progetto distribuito e scalabile che garantisce un accesso e una disponibilità eccellenti dei dati. Tuttavia, le aziende che utilizzano questa architettura devono far fronte a costi crescenti associati alla scalabilità per soddisfare il volume di dati in rapida crescita.

Splunk SmartStore con NetApp StorageGRID risolve questa sfida offrendo un nuovo modello di distribuzione in cui elaborazione e storage sono disaccoppiati. Questa soluzione sblocca inoltre una scalabilità e un'elasticità senza pari per gli ambienti Splunk Enterprise, consentendo ai clienti di scalare su siti singoli e multipli, riducendo al contempo i costi grazie alla scalabilità indipendente di elaborazione e archiviazione e aggiungendo livelli intelligenti all'archiviazione di oggetti S3 basata su cloud a costi contenuti.

La soluzione ottimizza la quantità di dati nell'archiviazione locale mantenendo al contempo le prestazioni di ricerca, consentendo di scalare elaborazione e archiviazione su richiesta. SmartStore valuta automaticamente i modelli di accesso ai dati per determinare quali dati devono essere accessibili per analisi in tempo reale e quali dati devono risiedere nell'archiviazione di oggetti S3 a basso costo.

Questo rapporto tecnico illustra i vantaggi che NetApp offre a una soluzione Splunk SmartStore, illustrando al contempo un framework per la progettazione e il dimensionamento di Splunk SmartStore nel tuo ambiente. Il risultato è una soluzione semplice, scalabile e resiliente che garantisce un TCO interessante. StorageGRID fornisce un archivio di oggetti scalabile e conveniente basato sul protocollo S3/API, noto anche come archiviazione remota, che consente alle organizzazioni di scalare la propria soluzione Splunk a un costo inferiore, aumentando al contempo la resilienza.

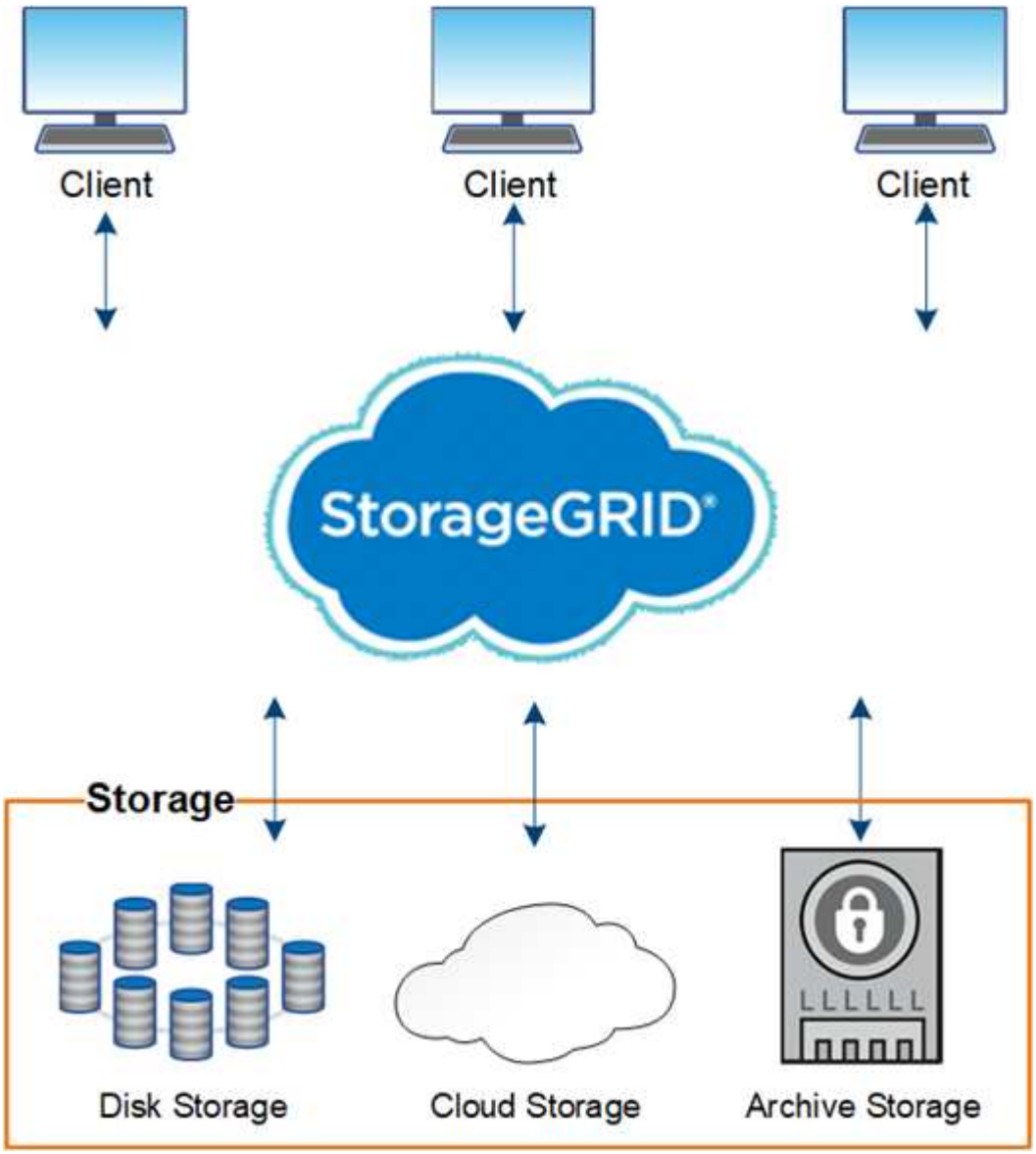
Splunk SmartStore definisce l'archiviazione di oggetti come archivi remoti o livelli di archiviazione remoti.

Informazioni su NetApp StorageGRID

NetApp StorageGRID è una soluzione di archiviazione di oggetti definita dal software per archivi di grandi dimensioni, repository multimediali e archivi di dati Web. Con StorageGRID, NetApp sfrutta due decenni di esperienza nella fornitura di soluzioni di innovazione e gestione dei dati leader del settore, aiutando al contempo le organizzazioni a gestire e massimizzare il valore delle loro informazioni sia in sede che in distribuzioni cloud pubbliche, private o ibride.

StorageGRID fornisce un archivio sicuro e durevole per dati non strutturati su larga scala. Le policy di gestione del ciclo di vita integrate e basate sui metadati ottimizzano la posizione dei dati durante tutto il loro ciclo di vita. I contenuti vengono posizionati nel posto giusto, al momento giusto e nel livello di archiviazione giusto per ridurre i costi. Grazie al singolo namespace è possibile accedere ai dati tramite un'unica chiamata, indipendentemente dalla posizione geografica dell'archiviazione StorageGRID . I clienti possono distribuire e gestire più istanze StorageGRID tra data center e nell'infrastruttura cloud.

Un sistema StorageGRID è composto da nodi eterogenei, ridondanti e distribuiti a livello globale, che possono essere integrati sia con applicazioni client esistenti che con quelle di nuova generazione.



IDC MarketScape ha recentemente nominato NetApp leader nell'ultimo rapporto, IDC MarketScape: Worldwide Object-Based Storage 2019 Vendor Assessment. Con quasi 20 anni di implementazione in ambito produttivo nei settori più esigenti, StorageGRID è un leader riconosciuto nel settore dei dati non strutturati.

Con StorageGRID puoi ottenere quanto segue:

- Distribuisci più istanze StorageGRID per accedere ai dati da qualsiasi posizione tra i data center e il cloud tramite un singolo namespace facilmente scalabile fino a centinaia di petabyte.
- Offrire flessibilità per distribuire e gestire centralmente le infrastrutture.
- Garantisce una durata senza pari con quindici-nove di resistenza sfruttando la codifica di cancellazione a strati (EC).
- Abilita più funzionalità multi-cloud ibride con integrazioni convalidate in Amazon S3 Glacier e Azure Blob.
- Soddisfa gli obblighi normativi e facilita la conformità tramite la conservazione dei dati a prova di manomissione, senza API proprietarie o vincoli con i fornitori.

Per ulteriori informazioni su come StorageGRID può aiutarti a risolvere i problemi più complessi di gestione dei dati non strutturati, consulta "Pagina iniziale di NetApp StorageGRID" .

Informazioni su Splunk Enterprise

Splunk Enterprise è una piattaforma che trasforma i dati in azioni concrete. I dati generati da varie fonti, come file di registro, siti Web, dispositivi, sensori e applicazioni, vengono inviati e analizzati dagli indicizzatori Splunk, consentendo di ricavare informazioni dettagliate dai dati. Potrebbe identificare violazioni dei dati, evidenziare tendenze di clienti e prodotti, trovare opportunità per ottimizzare l'infrastruttura o creare informazioni fruibili in un'ampia gamma di casi d'uso.

Informazioni su Splunk SmartStore

Splunk SmartStore amplia i vantaggi dell'architettura Splunk semplificandone al contempo la scalabilità in modo economicamente vantaggioso. La separazione delle risorse di elaborazione e di archiviazione si traduce in nodi indicizzatori ottimizzati per l'I/O con esigenze di archiviazione notevolmente ridotte, poiché memorizzano solo un sottoinsieme di dati nella cache. Non è necessario aggiungere ulteriore capacità di elaborazione o di archiviazione quando è necessaria solo una di queste risorse, il che consente di ottenere notevoli risparmi sui costi. È possibile utilizzare un archivio di oggetti basato su S3, conveniente e facilmente scalabile, che semplifica ulteriormente l'ambiente, riduce i costi e consente di gestire un set di dati più ampio.

Splunk SmartStore offre un valore significativo alle organizzazioni, tra cui:

- Riduzione dei costi di archiviazione spostando i dati caldi su un archivio di oggetti S3 ottimizzato in termini di costi
- Scalabilità senza soluzione di continuità mediante disaccoppiamento di storage e calcolo
- Semplificare la continuità aziendale sfruttando l'archiviazione cloud-native resiliente

Panoramica della soluzione

Questa pagina descrive i componenti utilizzati per completare questa soluzione, tra cui NetApp StorageGRID, Splunk Enterprise e Splunk SmartStore.

NetApp StorageGRID

NetApp StorageGRID è una piattaforma di storage di oggetti ad alte prestazioni e conveniente. Offre una gestione intelligente dei dati globali basata su policy, utilizzando un'architettura a griglia distribuita basata su nodi. Semplifica la gestione di petabyte di dati non strutturati e miliardi di oggetti attraverso il suo onnipresente spazio dei nomi degli oggetti globali combinato con sofisticate funzionalità di gestione dei dati. L'accesso agli oggetti tramite una singola chiamata si estende su più siti e semplifica le architetture ad alta disponibilità, garantendo al contempo un accesso continuo agli oggetti indipendentemente dalle interruzioni del sito o dell'infrastruttura.

La multitenancy consente di gestire in modo sicuro più applicazioni cloud e dati aziendali non strutturati all'interno della stessa griglia, aumentando il ROI e i casi d'uso di StorageGRID. È possibile creare più livelli di servizio con policy del ciclo di vita degli oggetti basate sui metadati, ottimizzando la durabilità, la protezione, le prestazioni e la località in più aree geografiche. Gli utenti possono adattare le policy e riallineare il panorama dei dati senza interruzioni man mano che cambiano le loro esigenze.

SmartStore sfrutta StorageGRID come livello di archiviazione remoto e consente ai clienti di implementare più siti distribuiti geograficamente per una disponibilità e una durabilità elevate, presentate come un singolo spazio dei nomi di oggetti. Ciò consente a Splunk SmartStore di sfruttare le elevate prestazioni, la capacità densa e la

capacità di scalare StorageGRID fino a centinaia di nodi su più siti fisici utilizzando un singolo URL per interagire con gli oggetti. Questo singolo URL consente inoltre che l'espansione, gli aggiornamenti e le riparazioni dello storage non interferiscano con le attività, anche al di fuori di un singolo sito. L'esclusivo motore di policy di gestione dei dati StorageGRID garantisce livelli ottimizzati di prestazioni e durabilità, nonché il rispetto dei requisiti di località dei dati.

Splunk Enterprise

Splunk, leader nella raccolta e nell'analisi di dati generati dalle macchine, contribuisce a semplificare e modernizzare l'IT attraverso le sue capacità di analisi operativa. Si estende anche ai casi d'uso di analisi aziendale, sicurezza e IoT. Lo storage è un fattore essenziale per il successo della distribuzione del software Splunk.

I dati generati dalle macchine rappresentano la tipologia di big data in più rapida crescita. Il formato è imprevedibile e proviene da molte fonti diverse, spesso a prezzi elevati e in grandi volumi. Queste caratteristiche del carico di lavoro vengono spesso definite "scarico digitale". Splunk SmartStore aiuta a dare un senso a questi dati e fornisce una suddivisione intelligente dei dati in livelli per il posizionamento ottimizzato dei dati attivi e passivi sul livello di archiviazione più conveniente.

Splunk SmartStore

Splunk SmartStore è una funzionalità di indicizzazione che utilizza l'archiviazione di oggetti (chiamata anche archiviazione remota o livelli di archiviazione remota) come StorageGRID per archiviare dati caldi tramite il protocollo S3.

Con l'aumento del volume di dati di un'implementazione, la domanda di spazio di archiviazione solitamente supera la domanda di risorse informatiche. SmartStore consente di gestire in modo conveniente le risorse di elaborazione e di archiviazione dell'indicizzatore, ridimensionando separatamente elaborazione e archiviazione.

SmartStore introduce un livello di archiviazione remoto, utilizzando il protocollo S3, e un gestore della cache. Queste funzionalità consentono ai dati di risiedere localmente su indicizzatori o su archivi remoti. Il gestore della cache, che risiede sull'indicizzatore, gestisce lo spostamento dei dati tra l'indicizzatore e il livello di archiviazione remoto. I dati vengono archiviati in bucket (caldi e tiepidi) insieme ai metadati dei bucket.

Con SmartStore puoi ridurre al minimo l'ingombro di archiviazione dell'indicizzatore e scegliere risorse di elaborazione ottimizzate per l'I/O, poiché la maggior parte dei dati risiede sul livello di archiviazione remoto. L'indicizzatore mantiene una cache locale, che rappresenta la quantità minima di dati necessaria per restituire i risultati richiesti e previsti. La cache locale contiene bucket attivi, copie di bucket attivi che partecipano a ricerche attive o recenti e metadati dei bucket.

Splunk SmartStore con StorageGRID consente ai clienti di scalare in modo incrementale l'ambiente con storage remoto ad alte prestazioni e conveniente, garantendo al contempo un elevato grado di elasticità alla soluzione complessiva. Ciò consente ai clienti di aggiungere qualsiasi componente (archiviazione a caldo e/o archiviazione S3 calda) in qualsiasi quantità e in qualsiasi momento, indipendentemente dal fatto che abbiano bisogno di più indicizzatori, di modificare la conservazione dei dati o di aumentare la velocità di acquisizione senza alcuna interruzione.

Vantaggi di questa soluzione

La soluzione consente di aggiungere risorse di elaborazione, hot storage o S3 per soddisfare la crescente domanda in termini di numero di utenti o velocità di acquisizione in distribuzioni singole e multi-sito.

- **Prestazione.** La combinazione di Splunk SmartStore e NetApp StorageGRID garantisce una rapida migrazione dei dati tra hot bucket e warm bucket utilizzando l'archiviazione di oggetti. StorageGRID accelera il processo di migrazione garantendo prestazioni rapide per carichi di lavoro di oggetti di grandi dimensioni.
- **Pronto per più siti.** L'architettura distribuita StorageGRID consente a Splunk SmartStore di estendere le distribuzioni su siti singoli e multipli tramite un unico namespace globale in cui è possibile accedere ai dati da qualsiasi sito, indipendentemente da dove si trovino.
- **Migliorata scalabilità.** Scala le risorse di storage in modo indipendente dalle risorse di elaborazione per soddisfare le esigenze e le richieste in continua evoluzione nel tuo ambiente Splunk, garantendo così un TCO migliore.
- **Capacità.** Gestisci i volumi in rapida crescita nella distribuzione Splunk con StorageGRID , scalando un singolo namespace a oltre 560 PB.
- **Disponibilità dei dati.** Ottimizza la disponibilità dei dati, le prestazioni, la distribuzione geografica, la conservazione, la protezione e i costi di archiviazione con policy basate sui metadati che possono essere adattate dinamicamente in base all'evoluzione del valore aziendale dei tuoi dati.

Aumenta le prestazioni con la cache SmartStore, un componente dell'indicizzatore che gestisce il trasferimento delle copie dei bucket tra l'archiviazione locale (hot) e quella remota (warm). Il dimensionamento Splunk per questa soluzione si basa su "linee guida fornite da Splunk" . La soluzione consente di aggiungere risorse di elaborazione, hot storage o S3 per soddisfare la crescente domanda in termini di numero di utenti o velocità di acquisizione in distribuzioni singole e multi-sito.

Architettura Splunk

Questa sezione descrive l'architettura di Splunk, comprese le definizioni chiave, le distribuzioni distribuite di Splunk, Splunk SmartStore, il flusso di dati, i requisiti hardware e software, i requisiti per siti singoli e multisito e così via.

Definizioni chiave

Le due tabelle successive elencano i componenti Splunk e NetApp utilizzati nella distribuzione di Splunk.

Questa tabella elenca i componenti hardware Splunk per la configurazione distribuita di Splunk Enterprise.

Componente Splunk	Compito
Indicizzatore	Repository per i dati di Splunk Enterprise
Spedizioniere universale	Responsabile dell'acquisizione dei dati e dell'inoltro dei dati agli indicizzatori
Testa di ricerca	L'interfaccia utente utilizzata per cercare dati negli indicizzatori
Maestro del cluster	Gestisce l'installazione Splunk di indicizzatori e testine di ricerca
Console di monitoraggio	Strumento di monitoraggio centralizzato utilizzato nell'intera distribuzione
Master di licenza	Il master delle licenze gestisce le licenze di Splunk Enterprise

Componente Splunk	Compito
Server di distribuzione	Aggiorna le configurazioni e distribuisce le app al componente di elaborazione
Componente di archiviazione	Compito
NetApp AFF	Storage all-flash utilizzato per gestire i dati di livello caldo. Noto anche come archiviazione locale.
NetApp StorageGRID	Archiviazione di oggetti S3 utilizzata per gestire i dati di livello caldo. Utilizzato da SmartStore per spostare i dati tra il livello caldo e quello caldo. Noto anche come archiviazione remota.

Questa tabella elenca i componenti dell'architettura di archiviazione Splunk.

Componente Splunk	Compito	Componente responsabile
Negozio intelligente	Fornisce agli indicizzatori la possibilità di suddividere i dati dall'archiviazione locale all'archiviazione degli oggetti.	Splunk
Caldo	Il punto di atterraggio in cui gli inoltratori universali inseriscono i dati appena scritti. L'archiviazione è scrivibile e i dati sono ricercabili. Questo livello di dati è in genere composto da SSD o HDD veloci.	ONTAP
Gestore della cache	Gestisce la cache locale dei dati indicizzati, recupera i dati caldi dall'archivio remoto quando si verifica una ricerca ed elimina dalla cache i dati utilizzati meno frequentemente.	Negozio intelligente
Caldo	I dati vengono trasferiti logicamente al bucket e rinominati prima dal livello caldo al livello caldo. I dati all'interno di questo livello sono protetti e, come nel livello caldo, possono essere composti da SSD o HDD di capacità maggiore. Sono supportati sia i backup incrementali che quelli completi utilizzando le comuni soluzioni di protezione dei dati.	StorageGRID

Distribuzioni distribuite di Splunk

Per supportare ambienti più grandi in cui i dati provengono da numerose macchine, è necessario elaborare grandi volumi di dati. Se molti utenti devono effettuare ricerche nei dati, è possibile scalare la distribuzione distribuendo le istanze di Splunk Enterprise su più macchine. Questo è noto come distribuzione distribuita.

In una tipica distribuzione distribuita, ogni istanza di Splunk Enterprise esegue un'attività specializzata e risiede su uno dei tre livelli di elaborazione corrispondenti alle principali funzioni di elaborazione.

Nella tabella seguente sono elencati i livelli di elaborazione di Splunk Enterprise.

Livello	Componente	Descrizione
Inserimento dati	Spedizioniere	Un forwarder consuma i dati e poi li inoltra a un gruppo di indicizzatori.
Indicizzazione	Indicizzatore	Un indicizzatore indicizza i dati in arrivo che solitamente riceve da un gruppo di inoltratori. L'indicizzatore trasforma i dati in eventi e memorizza gli eventi in un indice. L'indicizzatore ricerca anche i dati indicizzati in risposta alle richieste di ricerca provenienti da una testina di ricerca.
Gestione della ricerca	Testa di ricerca	Una testina di ricerca funge da risorsa centrale per la ricerca. Le teste di ricerca in un cluster sono intercambiabili e hanno accesso alle stesse ricerche, dashboard, oggetti di conoscenza e così via, da qualsiasi membro del cluster delle teste di ricerca.

Nella tabella seguente sono elencati i componenti importanti utilizzati in un ambiente Splunk Enterprise distribuito.

Componente	Descrizione	Responsabilità
Indice del cluster master	Coordina le attività e gli aggiornamenti di un cluster di indicizzatori	Gestione degli indici
Cluster di indice	Gruppo di indicizzatori Splunk Enterprise configurati per replicare i dati tra loro	Indicizzazione
Distributore della testina di ricerca	Gestisce la distribuzione e gli aggiornamenti al master del cluster	Gestione della testa di ricerca
Cluster di testine di ricerca	Gruppo di responsabili della ricerca che funge da risorsa centrale per la ricerca	Gestione della ricerca
Bilanciatori di carico	Utilizzato dai componenti in cluster per gestire la crescente domanda da parte di search head, indicizzatori e target S3 per distribuire il carico tra i componenti in cluster.	Gestione del carico per componenti raggruppati

Scopri i seguenti vantaggi delle distribuzioni distribuite di Splunk Enterprise:

- Accedere a fonti di dati diverse o disperse
- Fornire funzionalità per gestire le esigenze di dati per aziende di qualsiasi dimensione e complessità
- Ottieni un'elevata disponibilità e garantisci il ripristino di emergenza con la replica dei dati e la distribuzione multisito

Splunk SmartStore

SmartStore è una funzionalità di indicizzazione che consente agli archivi di oggetti remoti, come Amazon S3, di archiviare dati indicizzati. Con l'aumento del volume di dati di un'implementazione, la domanda di storage in genere supera la domanda di risorse di elaborazione. SmartStore consente di gestire in modo conveniente le risorse di archiviazione e di elaborazione dell'indicizzatore, ridimensionando tali risorse separatamente.

SmartStore introduce un livello di archiviazione remoto e un gestore della cache. Queste funzionalità consentono ai dati di risiedere localmente sugli indicizzatori o sul livello di archiviazione remoto. Il gestore della cache gestisce lo spostamento dei dati tra l'indicizzatore e il livello di archiviazione remoto, configurato sull'indicizzatore.

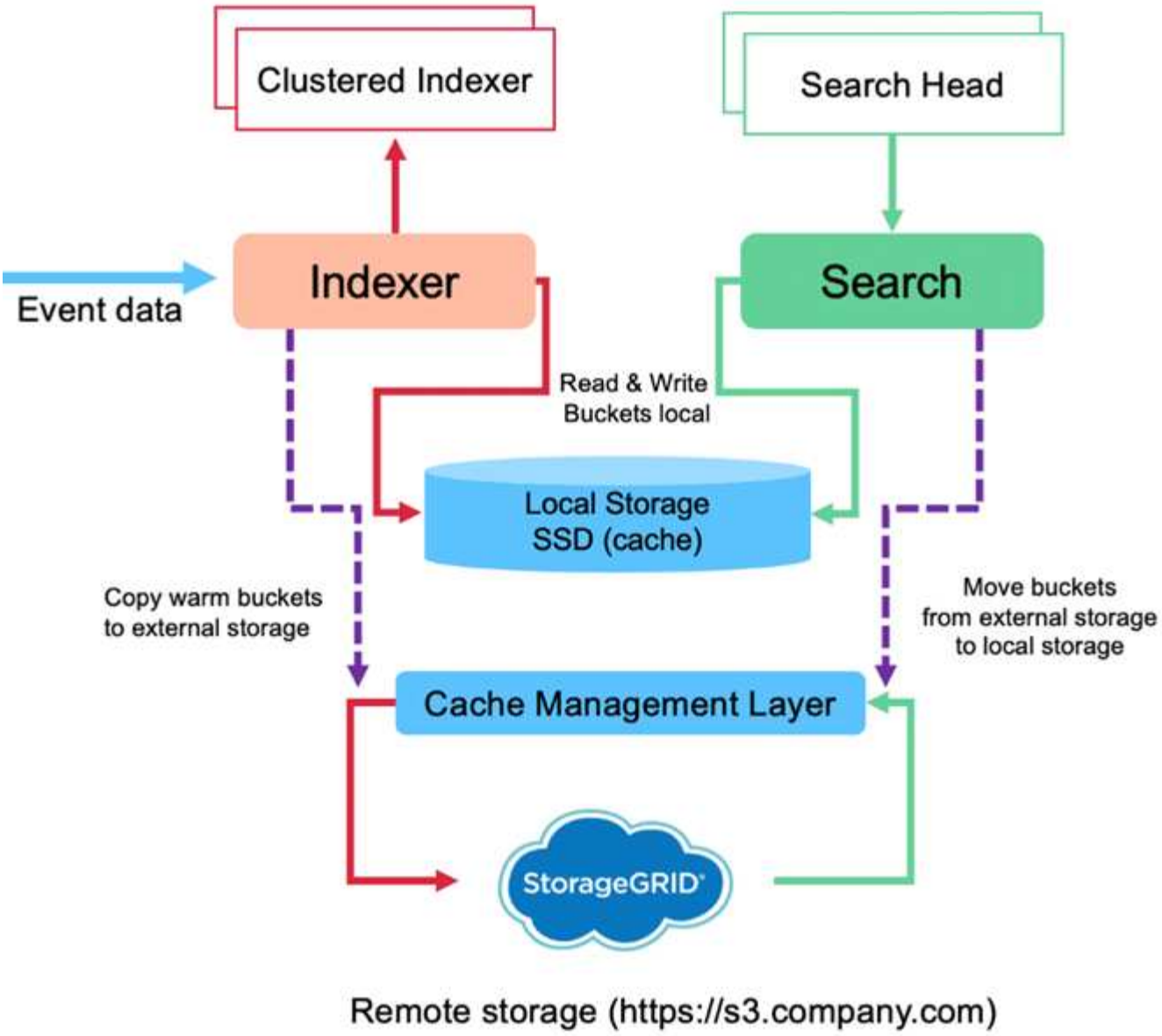
Con SmartStore puoi ridurre al minimo l'ingombro di archiviazione dell'indicizzatore e scegliere risorse di elaborazione ottimizzate per l'I/O. La maggior parte dei dati risiede nell'archiviazione remota. L'indicizzatore mantiene una cache locale che contiene una quantità minima di dati: hot bucket, copie di hot bucket che partecipano a ricerche attive o recenti e metadati dei bucket.

Flusso di dati di Splunk SmartStore

Quando i dati provenienti da varie fonti raggiungono gli indicizzatori, vengono indicizzati e salvati localmente in un hot bucket. L'indicizzatore replica anche i dati hot bucket sugli indicizzatori di destinazione. Finora, il flusso di dati è identico al flusso di dati per gli indici non SmartStore.

Quando il secchio caldo passa a quello caldo, il flusso di dati diverge. L'indicizzatore di origine copia il bucket caldo nell'archivio oggetti remoto (livello di archiviazione remoto) lasciando la copia esistente nella sua cache, perché le ricerche tendono a essere eseguite su dati indicizzati di recente. Tuttavia, gli indicizzatori di destinazione eliminano le proprie copie perché l'archivio remoto garantisce un'elevata disponibilità senza dover mantenere più copie locali. La copia master del bucket ora risiede nell'archivio remoto.

L'immagine seguente mostra il flusso di dati di Splunk SmartStore.



Il gestore della cache sull'indicizzatore è fondamentale per il flusso di dati SmartStore. Recupera copie dei bucket dall'archivio remoto secondo necessità per gestire le richieste di ricerca. Inoltre, rimuove dalla cache le copie più vecchie o meno ricercate dei bucket, perché la probabilità che partecipino alle ricerche diminuisce nel tempo.

Il compito del gestore della cache è ottimizzare l'uso della cache disponibile, garantendo al contempo che le ricerche abbiano accesso immediato ai bucket di cui hanno bisogno.

Requisiti software

Nella tabella seguente sono elencati i componenti software necessari per implementare la soluzione. I componenti software utilizzati in qualsiasi implementazione della soluzione potrebbero variare in base alle esigenze del cliente.

Famiglia di prodotti	Nome del prodotto	Versione del prodotto	Sistema operativo
NetApp StorageGRID	Archiviazione di oggetti StorageGRID	11,6	n / a
CentOS	CentOS	8,1	CentOS 7.x
Splunk Enterprise	Splunk Enterprise con SmartStore	8.0.3	CentOS 7.x

Requisiti per siti singoli e multisito

In un ambiente Splunk Enterprise (distribuzioni di medie e grandi dimensioni) in cui i dati hanno origine su più macchine e in cui molti utenti devono effettuare ricerche nei dati, è possibile scalare la distribuzione distribuendo le istanze di Splunk Enterprise su uno o più siti.

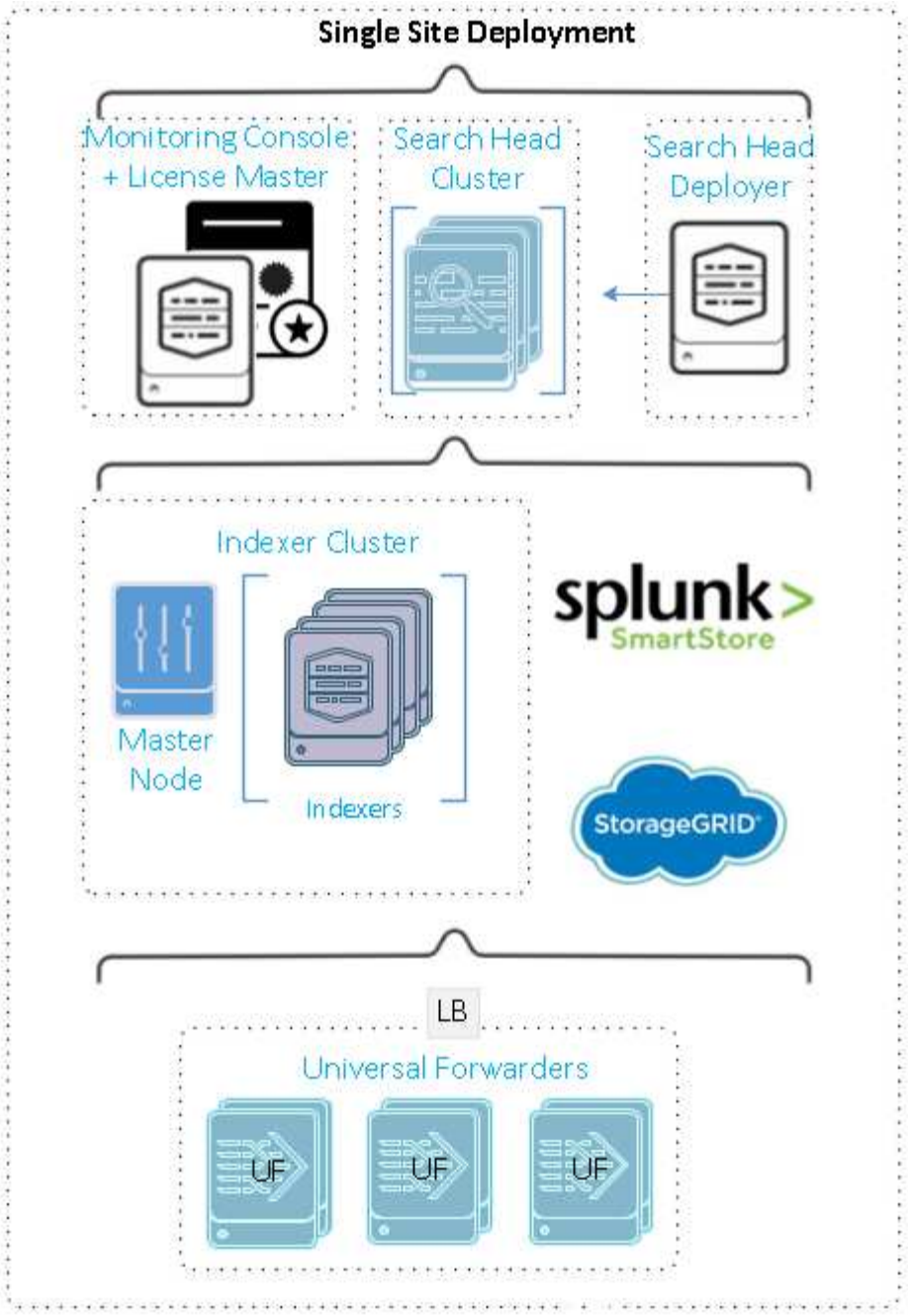
Scopri i seguenti vantaggi delle distribuzioni distribuite di Splunk Enterprise:

- Accedere a fonti di dati diverse o disperse
- Fornire funzionalità per gestire le esigenze di dati per aziende di qualsiasi dimensione e complessità
- Ottieni un'elevata disponibilità e garantisci il ripristino di emergenza con la replica dei dati e la distribuzione multisito

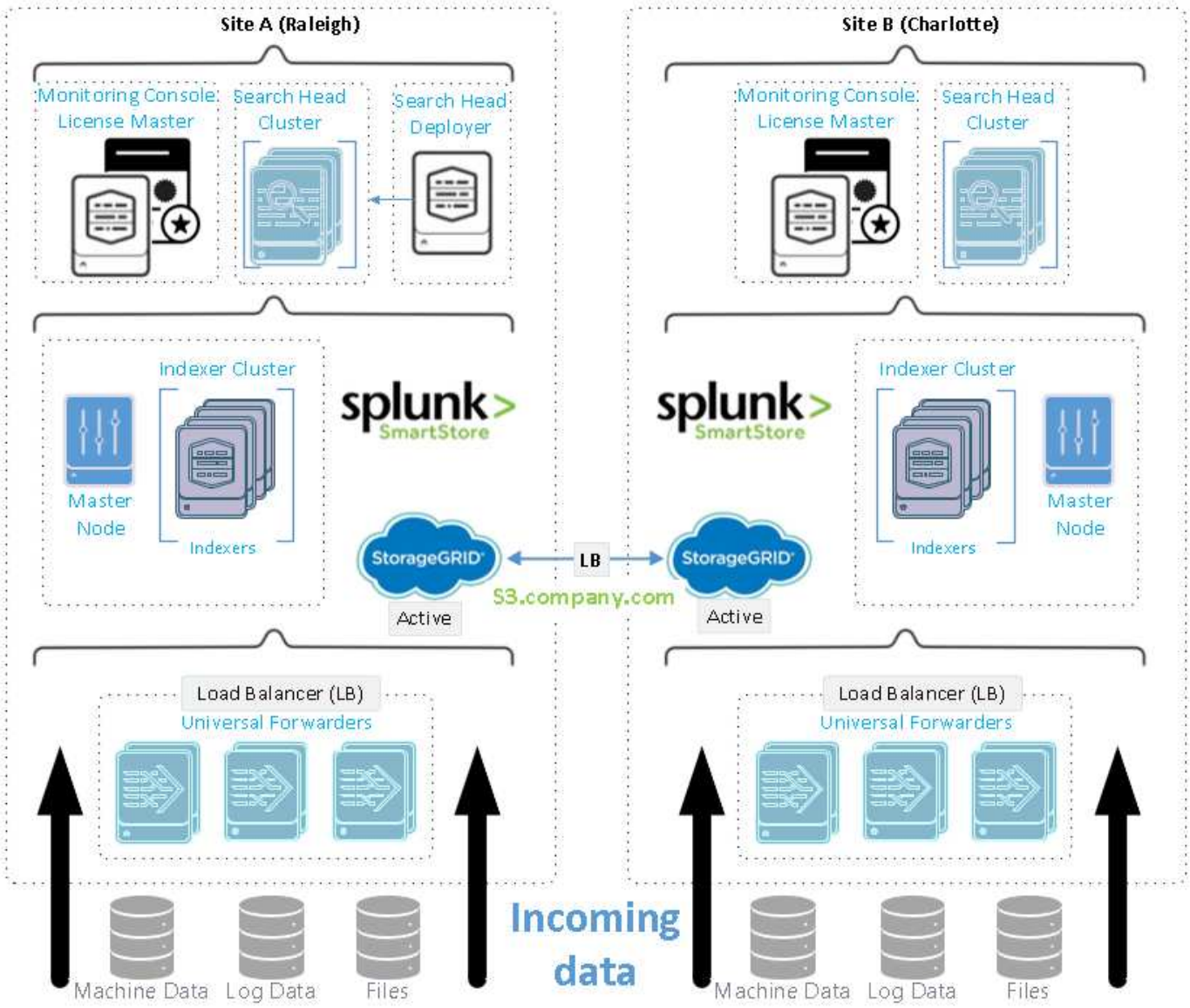
Nella tabella seguente sono elencati i componenti utilizzati in un ambiente Splunk Enterprise distribuito.

Componente	Descrizione	Responsabilità
Indice del cluster master	Coordina le attività e gli aggiornamenti di un cluster di indicizzatori	Gestione degli indici
Cluster di indice	Gruppo di indicizzatori Splunk Enterprise configurati per replicare i dati reciproci	Indicizzazione
Distributore della testina di ricerca	Gestisce la distribuzione e gli aggiornamenti al master del cluster	Gestione della testa di ricerca
Cluster di testine di ricerca	Gruppo di responsabili della ricerca che funge da risorsa centrale per la ricerca	Gestione della ricerca
Bilanciatori di carico	Utilizzato dai componenti in cluster per gestire la crescente domanda da parte di search head, indicizzatori e target S3 per distribuire il carico tra i componenti in cluster.	Gestione del carico per componenti raggruppati

Questa figura illustra un esempio di distribuzione su un singolo sito.



Questa figura illustra un esempio di distribuzione multisito.



Requisiti hardware

Le tabelle seguenti elencano il numero minimo di componenti hardware necessari per implementare la soluzione. I componenti hardware utilizzati in specifiche implementazioni della soluzione potrebbero variare in base alle esigenze del cliente.

> Indipendentemente dal fatto che Splunk SmartStore e StorageGRID siano stati distribuiti in un unico sito o in più siti, tutti i sistemi vengono gestiti da StorageGRID GRID Manager in un unico pannello di controllo. Per maggiori dettagli, consultare la sezione "Gestione semplice con Grid Manager".

Questa tabella elenca l'hardware utilizzato per un singolo sito.

Hardware	Quantità	Disco	Capacità utilizzabile	Nota
StorageGRID SG1000	1	n / a	n / a	Nodo di amministrazione e bilanciatore del carico
StorageGRID SG6060	4	x48, 8 TB (HDD NL-SAS)	1PB	Archiviazione remota

Questa tabella elenca l'hardware utilizzato per una configurazione multisito (per sito).

Hardware	Quantità	Disco	Capacità utilizzabile	Nota
StorageGRID SG1000	2	n / a	n / a	Nodo di amministrazione e bilanciatore del carico
StorageGRID SG6060	4	x48, 8 TB (HDD NL-SAS)	1PB	Archiviazione remota

Bilanciatore del carico NetApp StorageGRID : SG1000

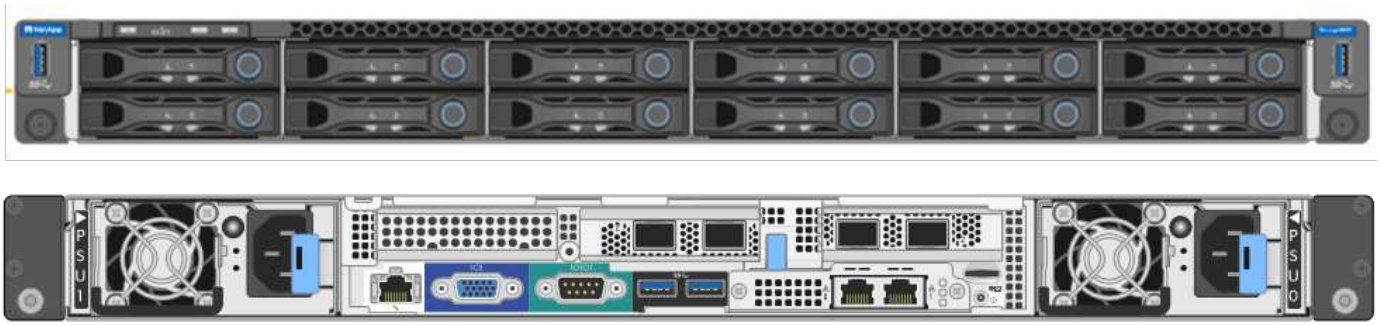
L'archiviazione di oggetti richiede l'uso di un bilanciatore del carico per presentare lo spazio dei nomi dell'archiviazione cloud. StorageGRID supporta bilanciatori di carico di terze parti di fornitori leader come F5 e Citrix, ma molti clienti scelgono il bilanciatore StorageGRID di livello aziendale per semplicità, resilienza e prestazioni elevate. Il bilanciatore del carico StorageGRID è disponibile come VM, container o appliance appositamente progettata.

StorageGRID SG1000 facilita l'uso di gruppi ad alta disponibilità (HA) e il bilanciamento del carico intelligente per le connessioni del percorso dati S3. Nessun altro sistema di archiviazione di oggetti on-premise fornisce un bilanciatore del carico personalizzato.

L'appliance SG1000 offre le seguenti funzionalità:

- Un bilanciatore del carico e, facoltativamente, funzioni di nodo di amministrazione per un sistema StorageGRID
- StorageGRID Appliance Installer per semplificare la distribuzione e la configurazione dei nodi
- Configurazione semplificata degli endpoint S3 e SSL
- Larghezza di banda dedicata (rispetto alla condivisione di un bilanciatore di carico di terze parti con altre applicazioni)
- Larghezza di banda Ethernet aggregata fino a 4 x 100 Gbps

L'immagine seguente mostra l'appliance SG1000 Gateway Services.

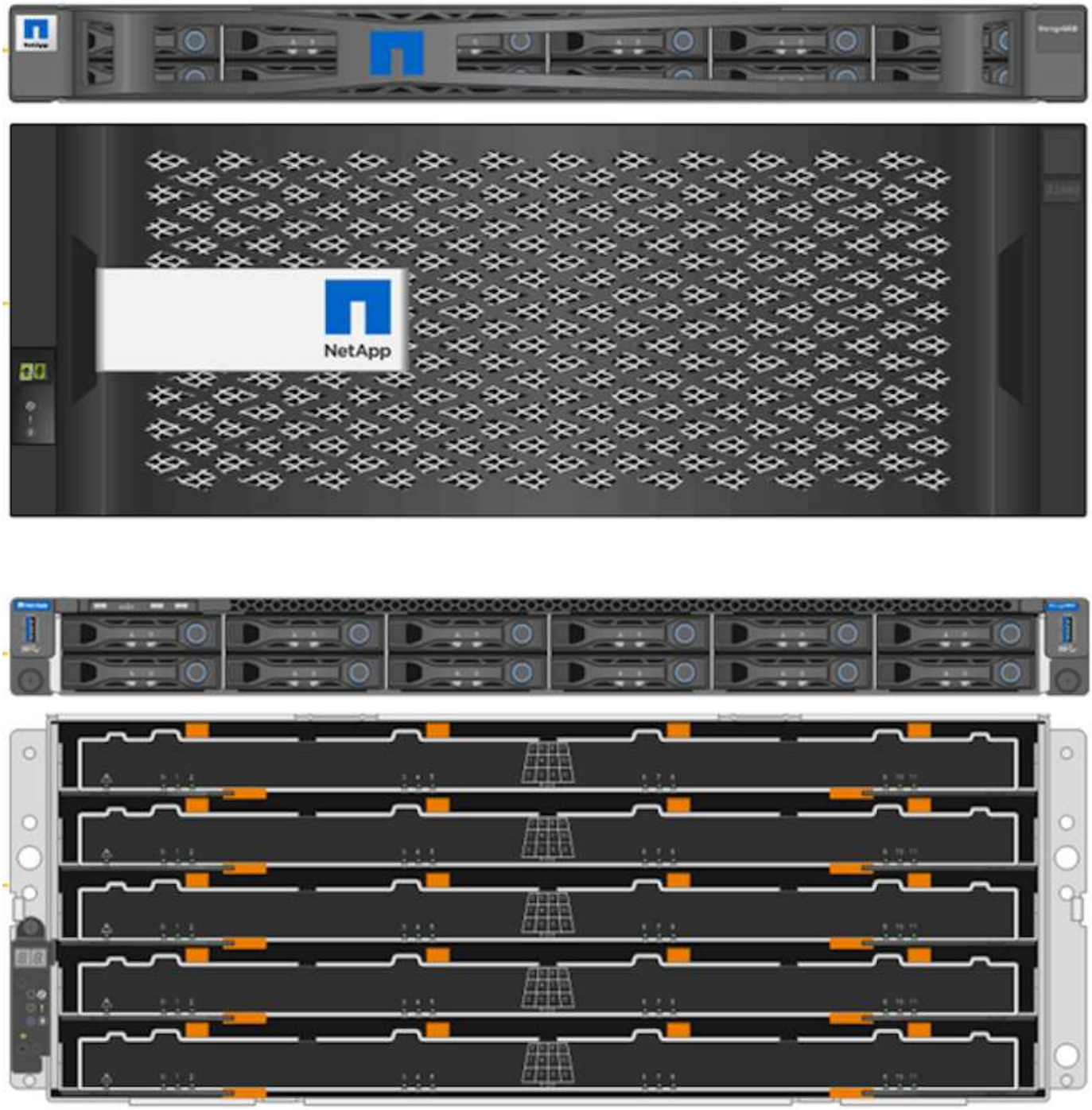


SG6060

L’appliance StorageGRID SG6060 include un controller di elaborazione (SG6060) e uno scaffale per controller di archiviazione (E-Series E2860) che contiene due controller di archiviazione e 60 unità. Questo apparecchio offre le seguenti caratteristiche:

- Scalabilità fino a 400 PB in un singolo namespace.
- Larghezza di banda Ethernet aggregata fino a 4x 25 Gbps.
- Include StorageGRID Appliance Installer per semplificare la distribuzione e la configurazione dei nodi.
- Ogni dispositivo SG6060 può avere uno o due ripiani di espansione aggiuntivi per un totale di 180 unità.
- Due controller E-Series E2800 (configurazione duplex) per fornire supporto failover del controller di archiviazione.
- Ripiano per unità a cinque cassetti che può contenere sessanta unità da 3,5 pollici (due unità a stato solido e 58 unità NL-SAS).

L’immagine seguente mostra l’appliance SG6060.



Progettazione Splunk

Nella tabella seguente è elencata la configurazione di Splunk per un singolo sito.

Componente Splunk	Compito	Quantità	Nuclei	Memoria	Sistema operativo
Spedizioniere universale	Responsabile dell'acquisizione dei dati e dell'inoltro dei dati agli indicizzatori	4	16 core	32 GB di RAM	CentOS 8.1
Indicizzatore	Gestisce i dati dell'utente	10	16 core	32 GB di RAM	CentOS 8.1
Testa di ricerca	Il front-end dell'utente cerca i dati negli indicizzatori	3	16 core	32 GB di RAM	CentOS 8.1
Distributore della testina di ricerca	Gestisce gli aggiornamenti per i cluster di testine di ricerca	1	16 core	32 GB di RAM	CentOS 8.1
Maestro del cluster	Gestisce l'installazione e gli indicizzatori di Splunk	1	16 core	32 GB di RAM	CentOS 8.1
Console di monitoraggio e master delle licenze	Esegue il monitoraggio centralizzato dell'intera distribuzione Splunk e gestisce le licenze Splunk	1	16 core	32 GB di RAM	CentOS 8.1

Le tabelle seguenti descrivono la configurazione di Splunk per configurazioni multisito.

Questa tabella elenca la configurazione Splunk per una configurazione multisito (sito A).

Componente Splunk	Compito	Quantità	Nuclei	Memoria	Sistema operativo
Spedizioniere universale	Responsabile dell'acquisizione dei dati e dell'inoltro dei dati agli indicizzatori.	4	16 core	32 GB di RAM	CentOS 8.1
Indicizzatore	Gestisce i dati dell'utente	10	16 core	32 GB di RAM	CentOS 8.1

Componente Splunk	Compito	Quantità	Nuclei	Memoria	Sistema operativo
Testa di ricerca	Il front-end dell'utente cerca i dati negli indicizzatori	3	16 core	32 GB di RAM	CentOS 8.1
Distributore della testina di ricerca	Gestisce gli aggiornamenti per i cluster di testine di ricerca	1	16 core	32 GB di RAM	CentOS 8.1
Maestro del cluster	Gestisce l'installazione e gli indicizzatori di Splunk	1	16 core	32 GB di RAM	CentOS 8.1
Console di monitoraggio e master delle licenze	Esegue il monitoraggio centralizzato dell'intera distribuzione Splunk e gestisce le licenze Splunk.	1	16 core	32 GB di RAM	CentOS 8.1

Questa tabella elenca la configurazione Splunk per una configurazione multisito (sito B).

Componente Splunk	Compito	Quantità	Nuclei	Memoria	Sistema operativo
Spedizioniere universale	Responsabile dell'acquisizione dei dati e dell'inoltro dei dati agli indicizzatori	4	16 core	32 GB di RAM	CentOS 8.1
Indicizzatore	Gestisce i dati dell'utente	10	16 core	32 GB di RAM	CentOS 8.1
Testa di ricerca	Il front-end dell'utente cerca i dati negli indicizzatori	3	16 core	32 GB di RAM	CentOS 8.1
Maestro del cluster	Gestisce l'installazione e gli indicizzatori di Splunk	1	16 core	32 GB di RAM	CentOS 8.1

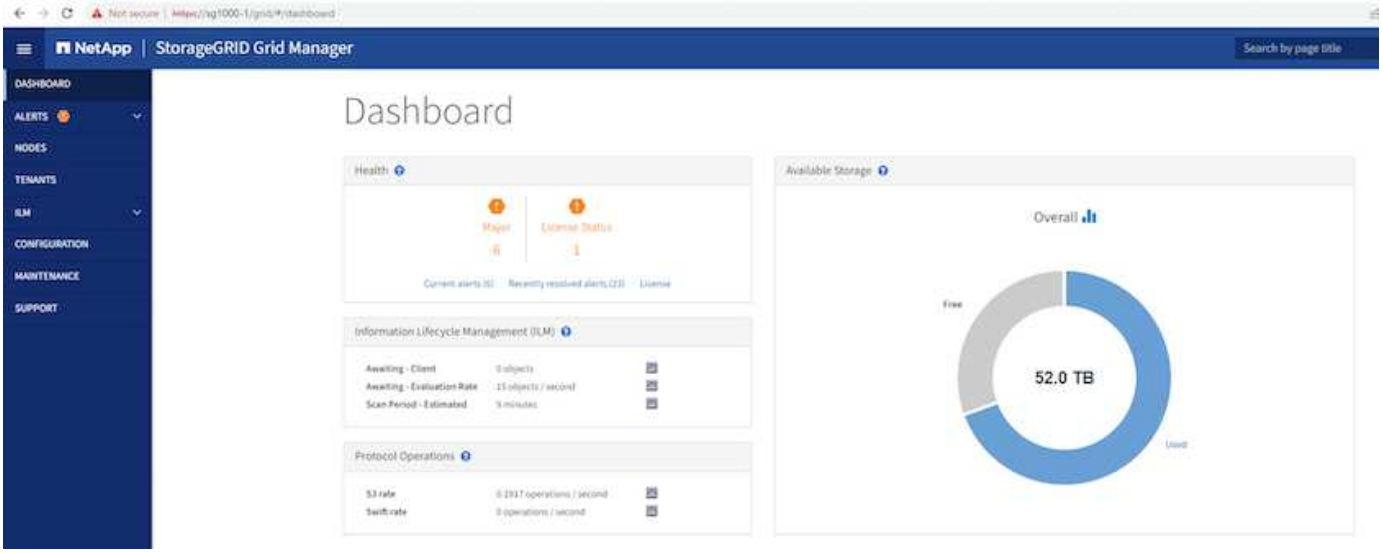
Componente Splunk	Compito	Quantità	Nuclei	Memoria	Sistema operativo
Console di monitoraggio e master delle licenze	Esegue il monitoraggio centralizzato dell'intera distribuzione Splunk e gestisce le licenze Splunk	1	16 core	32 GB di RAM	CentOS 8.1

Funzionalità flessibili StorageGRID per Splunk SmartStore

StorageGRID offre un'ampia gamma di funzionalità che gli utenti possono sfruttare e personalizzare in base al loro ambiente in continua evoluzione. Dall'implementazione al ridimensionamento di Splunk SmartStore, il tuo ambiente richiede un'adozione rapida dei cambiamenti e non deve interrompere Splunk. Le policy di gestione dati flessibili (ILM) e i classificatori del traffico (QoS) StorageGRID consentono di pianificare e adattare al proprio ambiente.

Gestione semplice con Grid Manager

Grid Manager è l'interfaccia grafica basata su browser che consente di configurare, gestire e monitorare il sistema StorageGRID in sedi distribuite a livello globale da un unico pannello di controllo, come mostrato nell'immagine seguente.



Eseguire le seguenti attività con l'interfaccia di Grid Manager:

- Gestisci repository di oggetti quali immagini, video e record distribuiti a livello globale, su scala petabyte.
- Monitorare i nodi e i servizi della griglia per garantire la disponibilità degli oggetti.
- Gestire il posizionamento dei dati degli oggetti nel tempo utilizzando le regole di gestione del ciclo di vita delle informazioni (ILM). Queste regole stabiliscono cosa accade ai dati di un oggetto dopo che sono stati acquisiti, come vengono protetti dalla perdita, dove vengono archiviati i dati dell'oggetto e per quanto

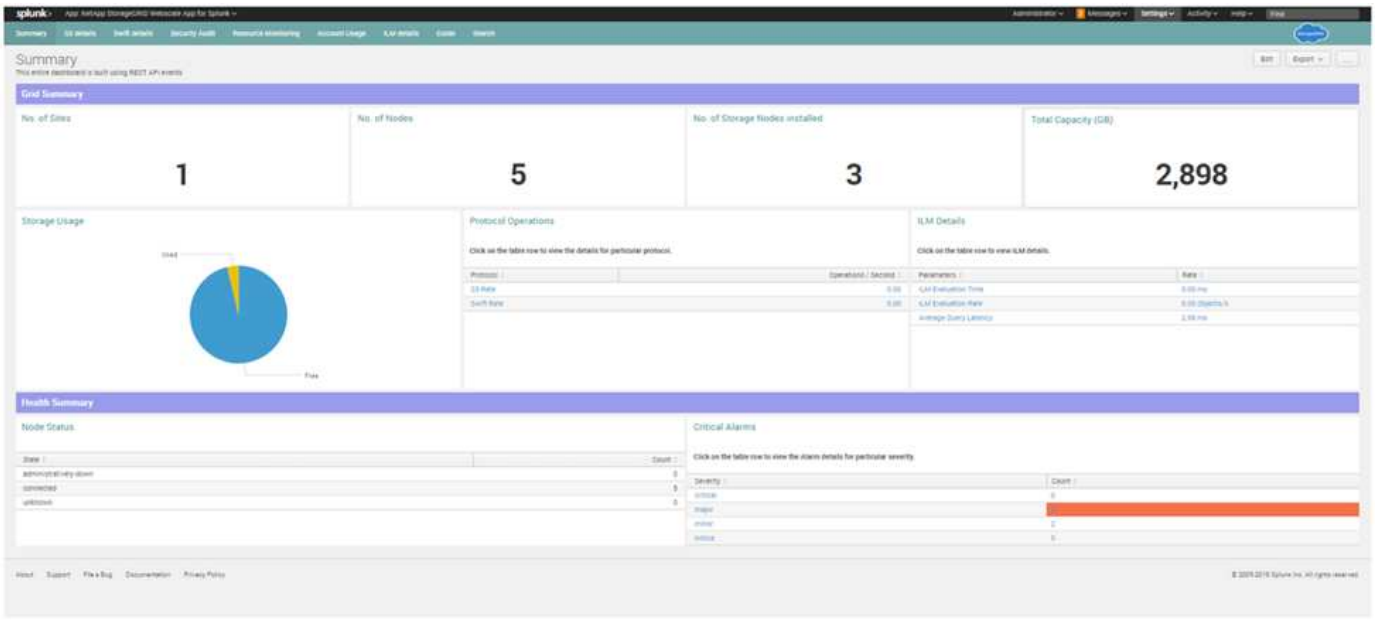
tempo.

- Monitorare le transazioni, le prestazioni e le operazioni all'interno del sistema.

Applicazione NetApp StorageGRID per Splunk

L'app NetApp StorageGRID per Splunk è un'applicazione specifica per Splunk Enterprise. Questa app funziona insieme al componente aggiuntivo NetApp StorageGRID per Splunk. Fornisce visibilità sullo stato StorageGRID , informazioni sull'utilizzo dell'account, dettagli di controllo della sicurezza, utilizzo e monitoraggio delle risorse e così via.

L'immagine seguente mostra l'app StorageGRID per Splunk.



Politiche ILM

StorageGRID dispone di policy di gestione dei dati flessibili che includono la conservazione di più copie degli oggetti e l'utilizzo di schemi EC (erasure coding) come 2+1 e 4+2 (e molti altri) per archiviare gli oggetti in base a requisiti specifici di prestazioni e protezione dei dati. Poiché i carichi di lavoro e i requisiti cambiano nel tempo, è normale che anche le policy ILM debbano cambiare nel tempo. La modifica delle policy ILM è una funzionalità fondamentale che consente ai clienti StorageGRID di adattarsi in modo rapido e semplice al loro ambiente in continua evoluzione.

Prestazione

StorageGRID aumenta le prestazioni aggiungendo più nodi, che possono essere VM, bare metal o appliance appositamente realizzate come SG5712, SG5760, SG6060 o SGF6024. Nei nostri test abbiamo superato i requisiti di prestazioni chiave di SmartStore con una griglia minima a tre nodi utilizzando l'appliance SG6060. Man mano che i clienti ampliano la propria infrastruttura Splunk con indicizzatori aggiuntivi, possono aggiungere più nodi di archiviazione per aumentare prestazioni e capacità.

Configurazione del bilanciatore del carico e degli endpoint

I nodi amministrativi in StorageGRID forniscono l'interfaccia utente (UI) di Grid Manager e l'endpoint API REST per visualizzare, configurare e gestire il sistema StorageGRID , nonché registri di controllo per monitorare l'attività del sistema. Per fornire un endpoint S3 ad alta disponibilità per l'archiviazione remota Splunk

SmartStore, abbiamo implementato il bilanciatore del carico StorageGRID , che viene eseguito come servizio sui nodi di amministrazione e sui nodi gateway. Inoltre, il bilanciatore del carico gestisce anche il traffico locale e comunica con il GSLB (Global Server Load Balancing) per facilitare il ripristino in caso di emergenza.

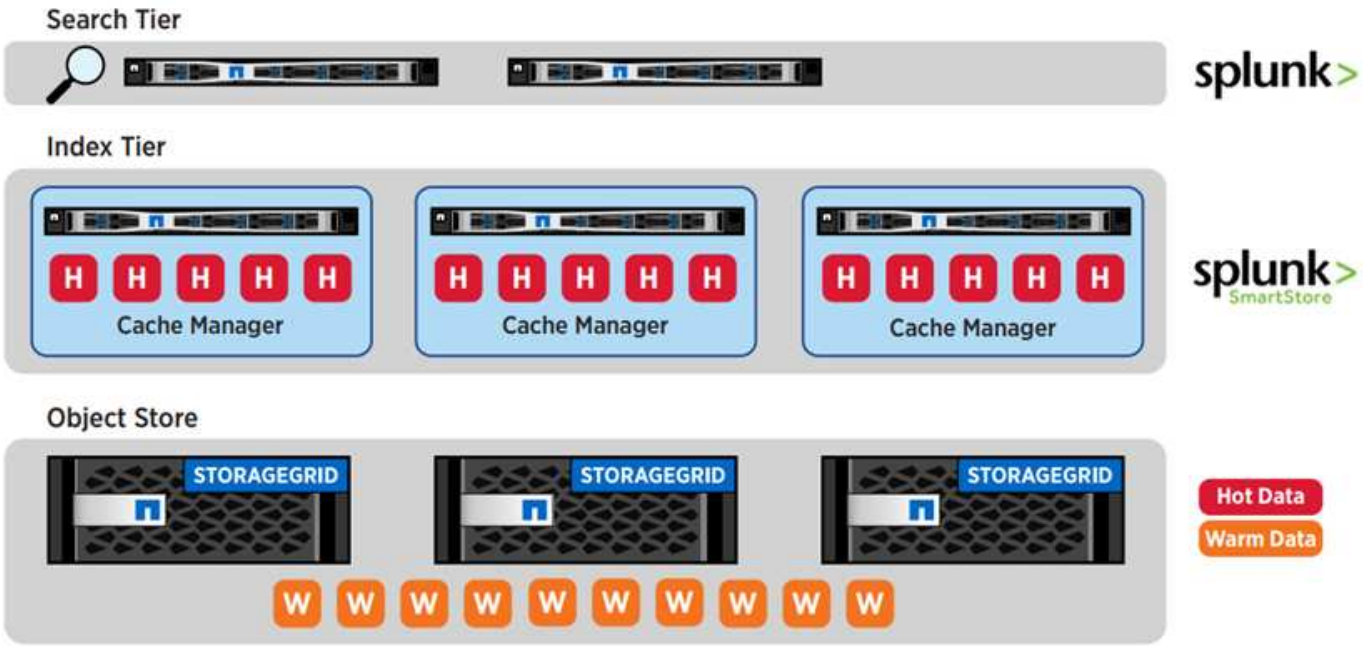
Per migliorare ulteriormente la configurazione degli endpoint, StorageGRID fornisce criteri di classificazione del traffico integrati nel nodo di amministrazione, consente di monitorare il traffico del carico di lavoro e di applicare vari limiti di qualità del servizio (QoS) ai carichi di lavoro. I criteri di classificazione del traffico vengono applicati agli endpoint del servizio StorageGRID Load Balancer per i nodi gateway e i nodi amministrativi. Queste politiche possono aiutare a limitare e monitorare il traffico.

Tiering intelligente e risparmio sui costi

Man mano che i clienti si rendono conto della potenza e della semplicità di utilizzo dell'analisi dei dati di Splunk, è naturale che vogliano indicizzare una quantità di dati sempre maggiore. Con l'aumentare della quantità di dati, aumenta anche l'infrastruttura di elaborazione e archiviazione necessaria per gestirli. Poiché i dati più vecchi vengono consultati meno frequentemente, impegnare la stessa quantità di risorse di elaborazione e consumare costosi archivi primari diventa sempre più inefficiente. Per operare su larga scala, i clienti traggono vantaggio dallo spostamento dei dati "caldi" a un livello più conveniente, liberando risorse di elaborazione e di storage primario per i dati "caldi".

Splunk SmartStore con StorageGRID offre alle organizzazioni una soluzione scalabile, performante e conveniente. Poiché SmartStore è consapevole dei dati, valuta automaticamente i modelli di accesso ai dati per determinare quali dati devono essere accessibili per analisi in tempo reale (dati attivi) e quali dati devono risiedere in un archivio a lungo termine a basso costo (dati attivi). SmartStore utilizza in modo dinamico e intelligente l'API AWS S3, standard del settore, inserendo i dati nello storage S3 fornito da StorageGRID. L'architettura flessibile e scalabile di StorageGRID consente al livello di dati caldi di crescere in modo economicamente vantaggioso in base alle necessità. L'architettura basata su nodi di StorageGRID garantisce che i requisiti di prestazioni e costi siano soddisfatti in modo ottimale.

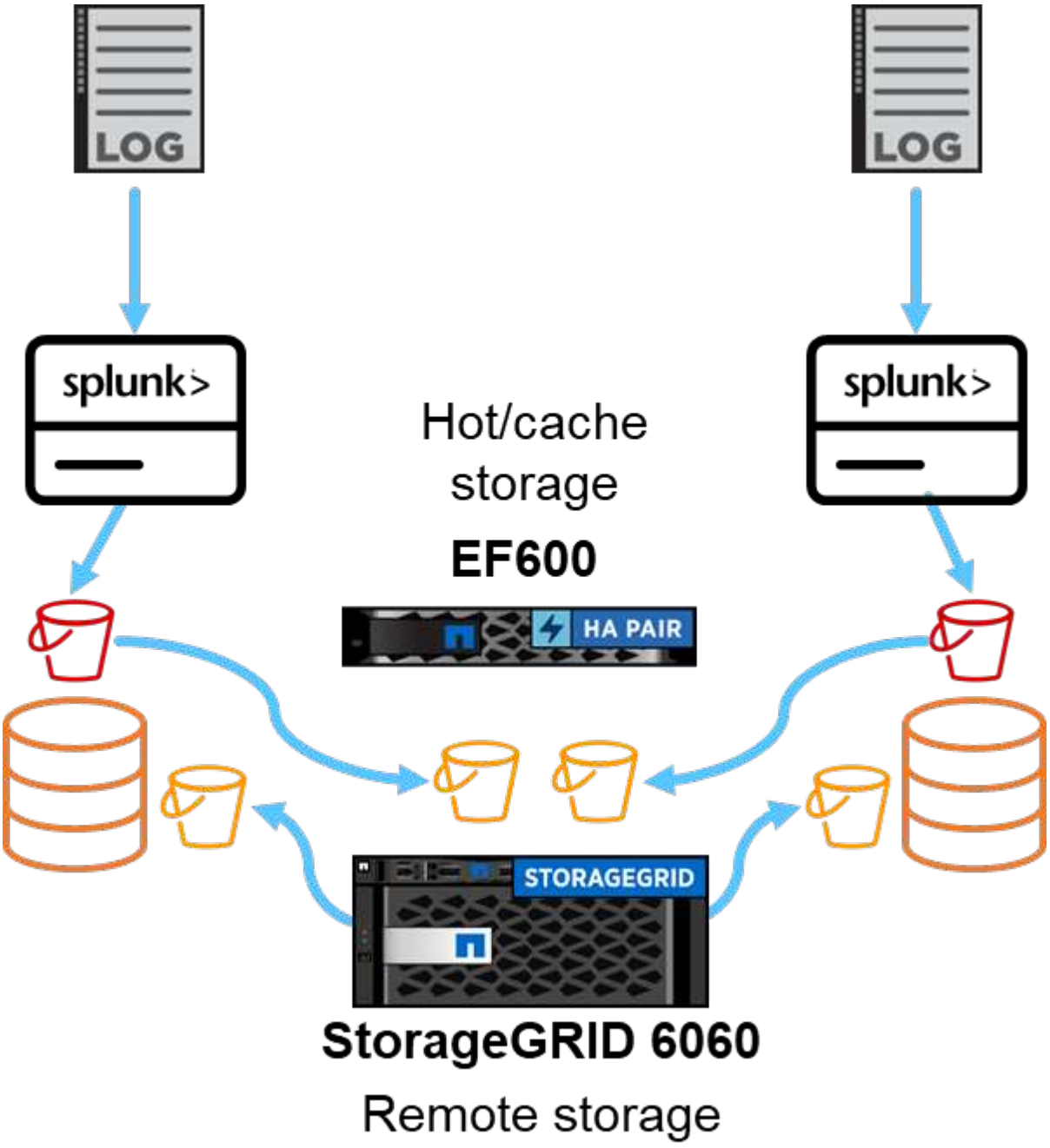
L'immagine seguente illustra la suddivisione in livelli di Splunk e StorageGRID .



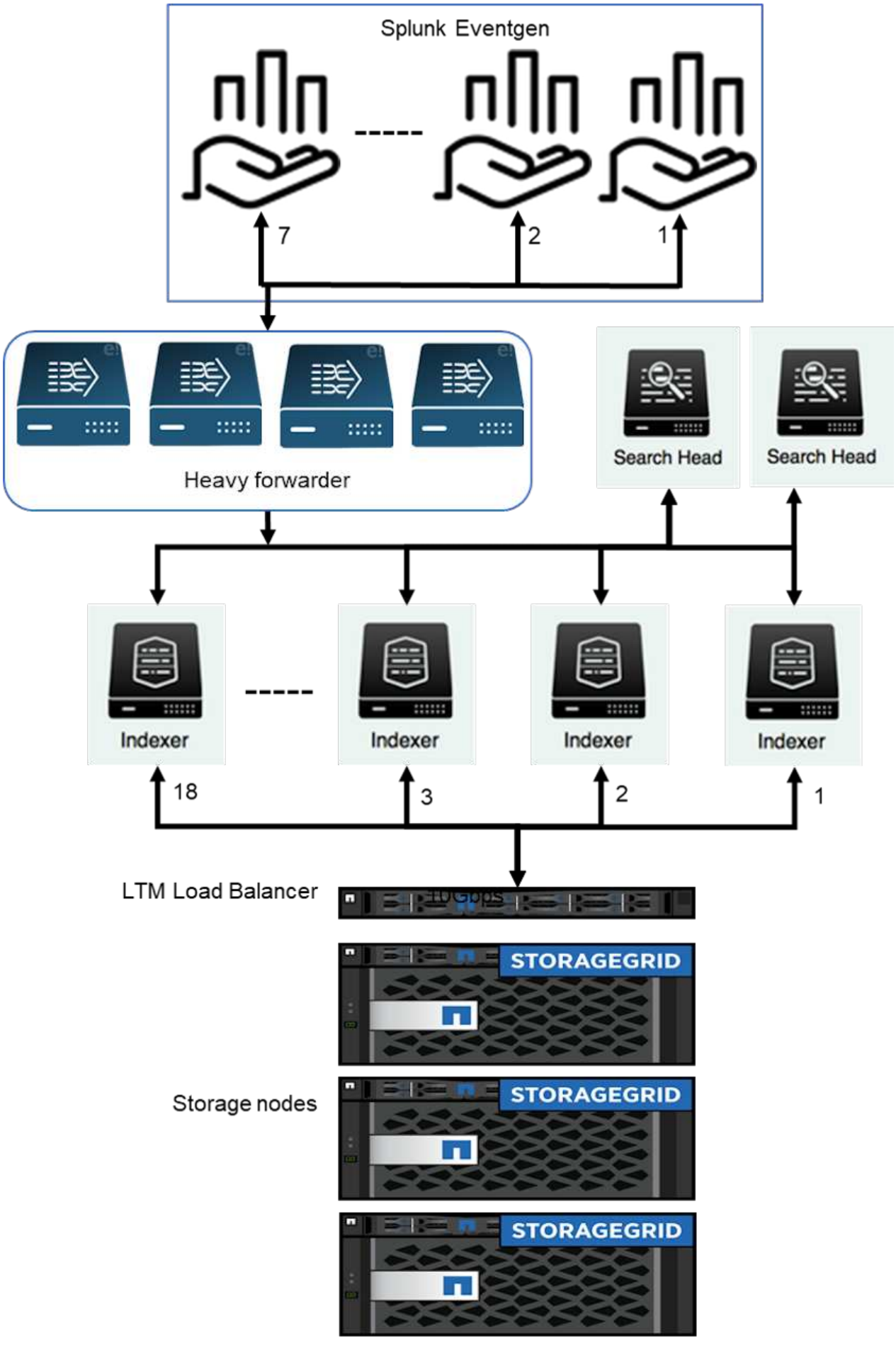
La combinazione leader del settore di Splunk SmartStore con NetApp StorageGRID offre i vantaggi di un'architettura disaccoppiata tramite una soluzione full-stack.

Prestazioni SmartStore a sito singolo

Questa sezione descrive le prestazioni di Splunk SmartStore su un controller NetApp StorageGRID . Splunk SmartStore sposta i dati caldi su un archivio remoto, che in questo caso è l'archivio di oggetti StorageGRID nella convalida delle prestazioni.



Abbiamo utilizzato EF600 per l'archiviazione hot/cache e StorageGRID 6060 per l'archiviazione remota. Per la convalida delle prestazioni abbiamo utilizzato la seguente architettura. Abbiamo utilizzato due search head, quattro heavy forwarder per inoltrare i dati agli indicizzatori, sette Splunk Event Generator (Eventgen) per generare i dati in tempo reale e 18 indicizzatori per archiviare i dati.


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

Configurazione

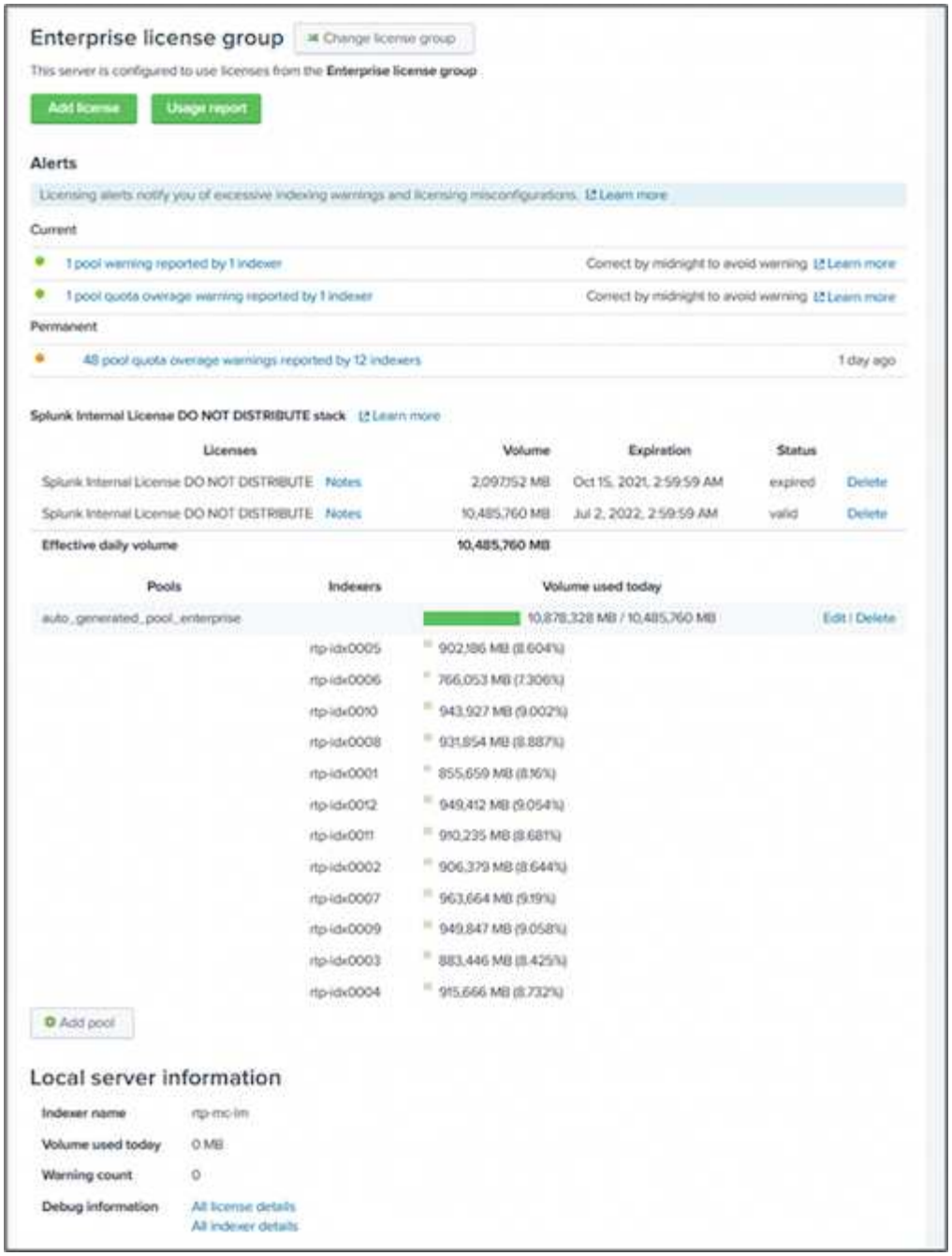
Questa tabella elenca l'hardware utilizzato per la convalida delle prestazioni di SmartStorage.

Componente Splunk	Compito	Quantità	Nuclei	Memoria	Sistema operativo
Spedizioniere pesante	Responsabile dell'acquisizione dei dati e dell'inoltro dei dati agli indicizzatori	4	16 core	32 GB di RAM	SLITTA 15 SP2
Indicizzatore	Gestisce i dati dell'utente	18	16 core	32 GB di RAM	SLITTA 15 SP2
Testa di ricerca	L'interfaccia utente cerca i dati negli indicizzatori	2	16 core	32 GB di RAM	SLITTA 15 SP2
Distributore della testina di ricerca	Gestisce gli aggiornamenti per i cluster di testine di ricerca	1	16 core	32 GB di RAM	SLITTA 15 SP2
Maestro del cluster	Gestisce l'installazione e gli indicizzatori di Splunk	1	16 core	32 GB di RAM	SLITTA 15 SP2
Console di monitoraggio e master delle licenze	Esegue il monitoraggio centralizzato dell'intera distribuzione Splunk e gestisce le licenze Splunk	1	16 core	32 GB di RAM	SLITTA 15 SP2

Convalida delle prestazioni del negozio remoto SmartStore

In questa convalida delle prestazioni, abbiamo configurato la cache SmartStore nell'archiviazione locale su tutti gli indicizzatori per 10 giorni di dati. Abbiamo abilitato il `maxDataSize=auto` (dimensione bucket da 750 MB) nel gestore cluster Splunk e ho inviato le modifiche a tutti gli indicizzatori. Per misurare le prestazioni di caricamento, abbiamo acquisito 10 TB al giorno per 10 giorni e abbiamo trasferito tutti i bucket attivi in modalità riscaldamento contemporaneamente, catturando il picco e la velocità effettiva media per istanza e per l'intera distribuzione dalla dashboard della SmartStore Monitoring Console.

Questa immagine mostra i dati acquisiti in un giorno.



Abbiamo eseguito il seguente comando dal cluster master (il nome dell'indice è `eventgen-test`). Abbiamo quindi registrato il picco e la velocità media di caricamento per istanza e per l'intera distribuzione tramite le dashboard della SmartStore Monitoring Console.

```
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
```

Il master del cluster dispone di autenticazione senza password per tutti gli indicizzatori (rtp-idx0001…rtp-idx0018).

Per misurare le prestazioni di download, abbiamo rimosso tutti i dati dalla cache eseguendo due volte l'interfaccia a riga di comando evict utilizzando il seguente comando.

Abbiamo eseguito il seguente comando dal cluster master ed eseguito la ricerca dalla testina di ricerca su 10 giorni di dati dall'archivio remoto da StorageGRID. Abbiamo quindi registrato il picco e la velocità media di caricamento per istanza e per l'intera distribuzione tramite le dashboard della SmartStore Monitoring Console.

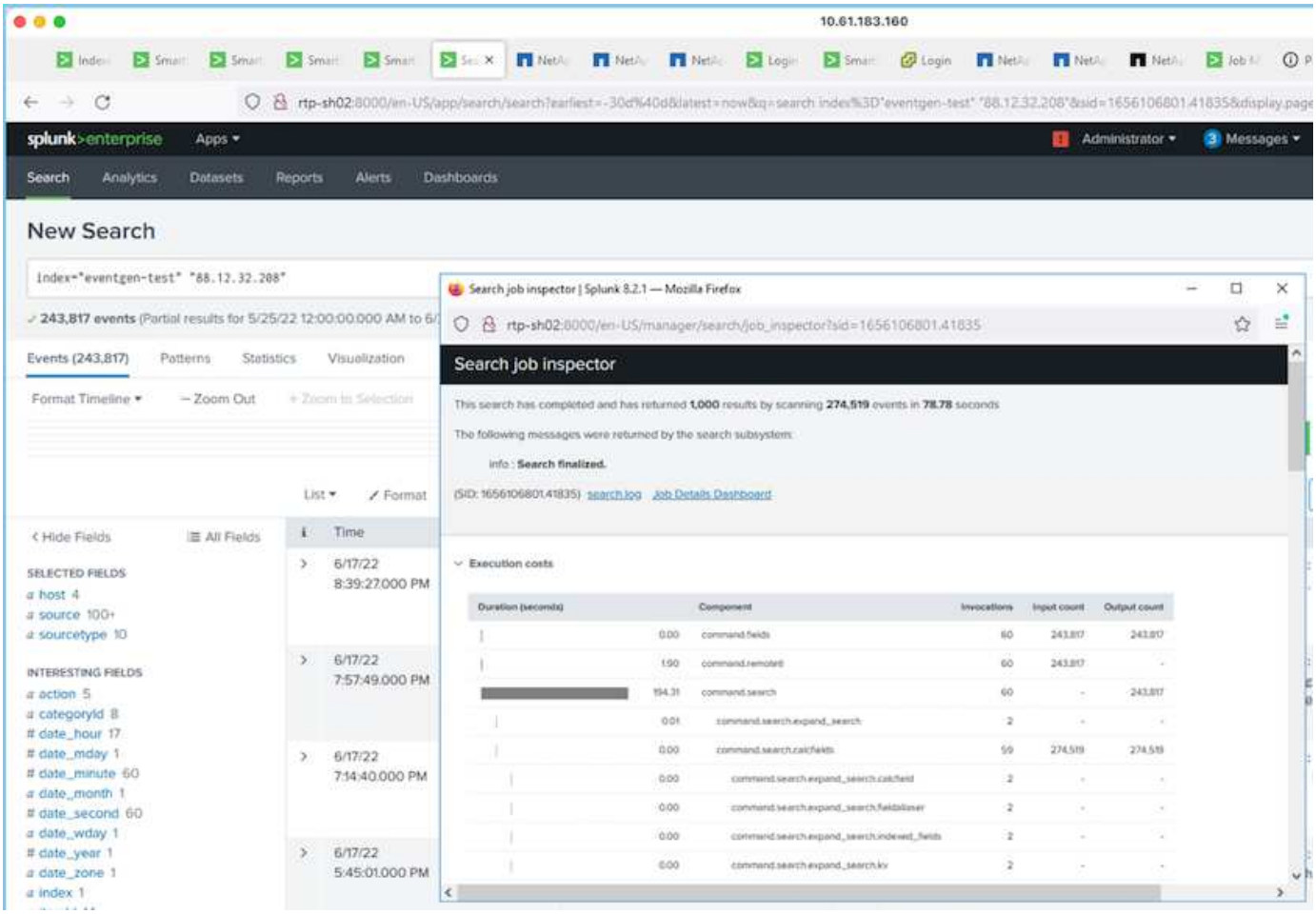
```
for i in rtp-idx0001 rtp-idx0002 rtp-idx0003 rtp-idx0004 rtp-idx0005 rtp-
idx0006 rtp-idx0007 rtp-idx0008 rtp-idx0009 rtp-idx0010 rtp-idx0011 rtp-
idx0012 rtp-idx0013 rtp-idx0014 rtp-idx0015 rtp-idx0016 rtp-idx0017 rtp-
idx0018 ; do  ssh $i " hostname;  date; /opt/splunk/bin/splunk _internal
call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

Le configurazioni dell'indicizzatore sono state trasferite dal master del cluster SmartStore. Il master del cluster aveva la seguente configurazione per l'indicizzatore.

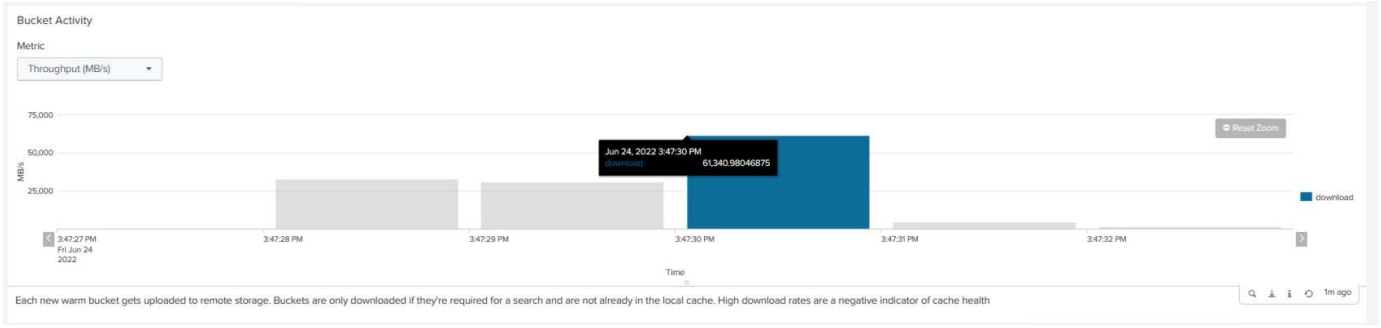
```
Rtp-cm01:~ # cat /opt/splunk/etc/master-apps/_cluster/local/indexes.conf
[default]
maxDataSize = auto
#defaultDatabase = eventgen-basic
defaultDatabase = eventgen-test
hotlist_recency_secs = 864000
repFactor = auto
[volume:remote_store]
storageType = remote
path = s3://smartstore2
remote.s3.access_key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

Abbiamo eseguito la seguente query di ricerca sulla testina di ricerca per raccogliere la matrice delle prestazioni.



Abbiamo raccolto le informazioni sulle prestazioni dal cluster master. La prestazione massima è stata di 61,34 GBps.

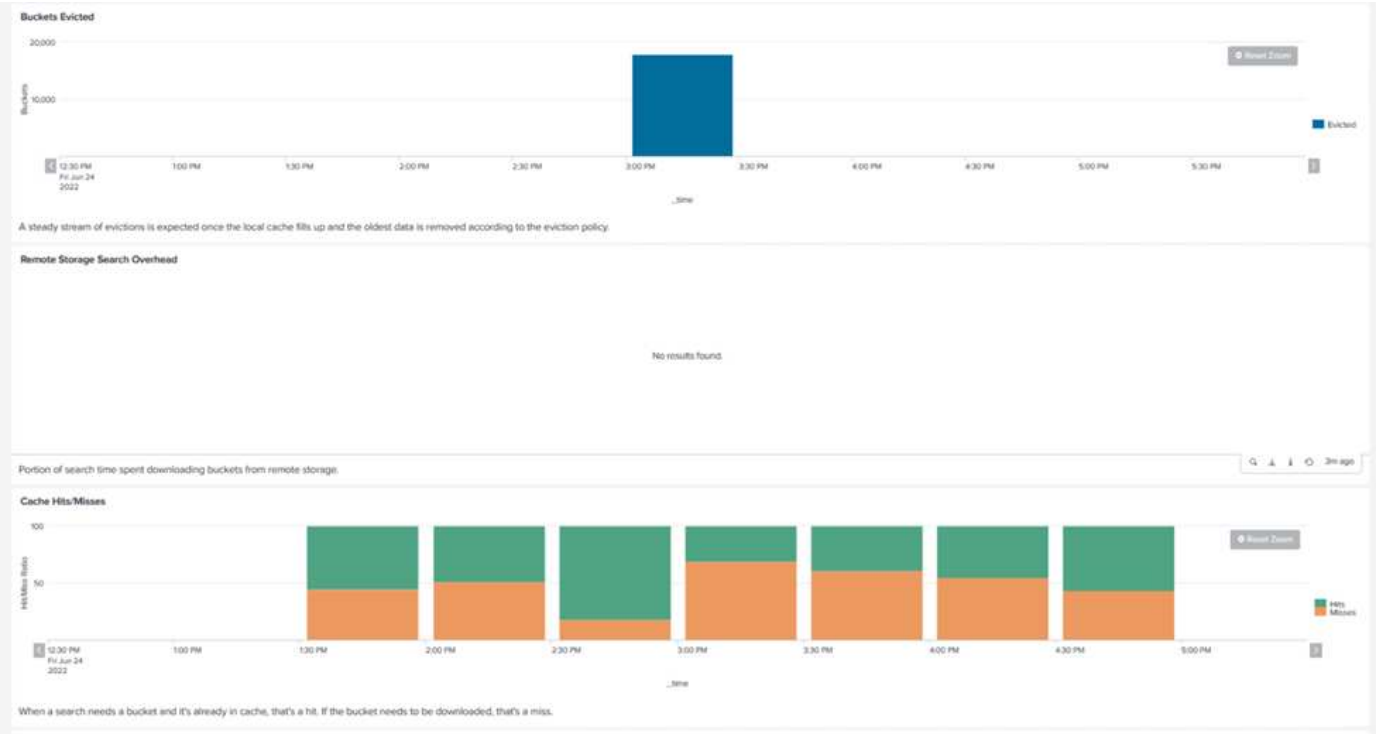


La prestazione media è stata di circa 29 GBps.

Prestazioni StorageGRID

Le prestazioni di SmartStore si basano sulla ricerca di modelli e stringhe specifici da grandi quantità di dati. In questa convalida, gli eventi vengono generati utilizzando "Eventgen" su uno specifico indice Splunk (eventgen-test) tramite la testina di ricerca e la richiesta viene inviata a StorageGRID per la maggior parte delle query. L'immagine seguente mostra i risultati positivi e negativi dei dati della query. I dati relativi agli hit provengono dal disco locale, mentre i dati relativi agli miss provengono dal controller StorageGRID .

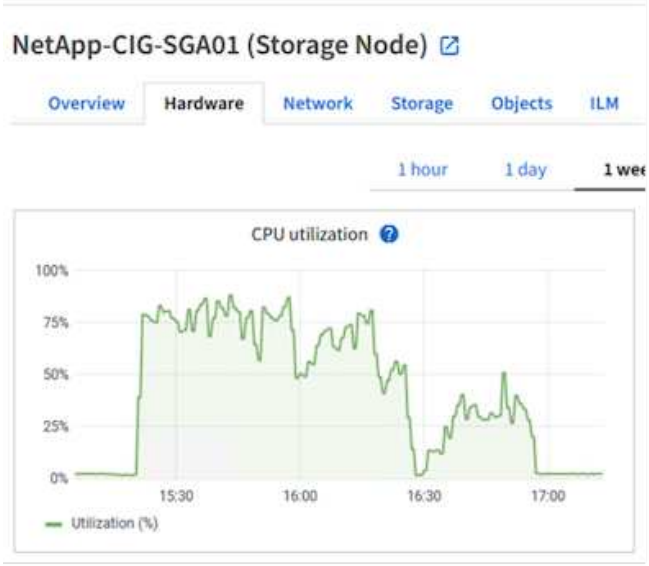
Il colore verde mostra i dati dei successi, mentre il colore arancione mostra i dati dei fallimenti.

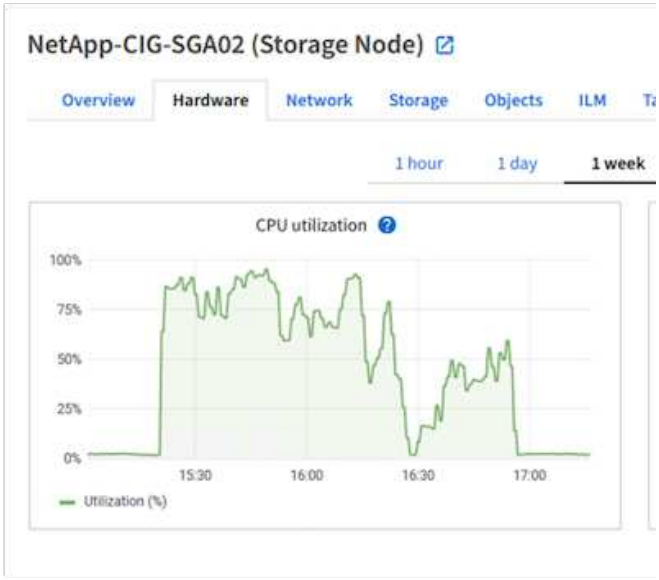


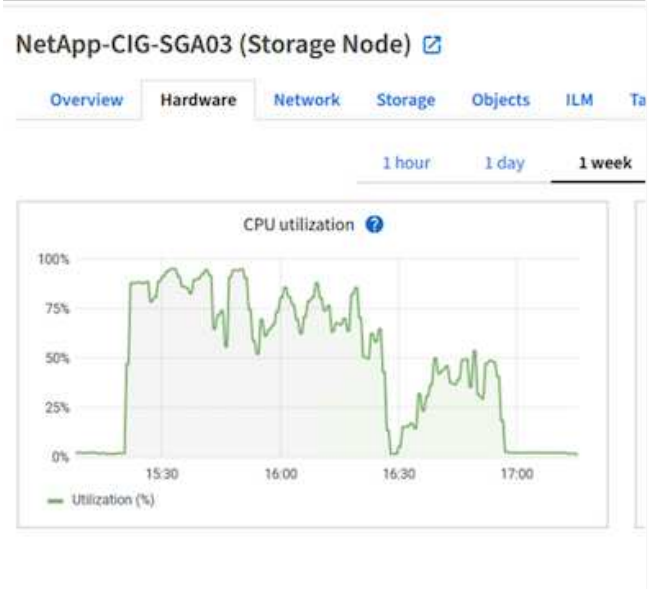
Quando viene eseguita la query per la ricerca su StorageGRID, il tempo di recupero S3 da StorageGRID viene mostrato nell'immagine seguente.

Utilizzo dell'hardware StorageGRID

L'istanza StorageGRID ha un bilanciatore del carico e tre controller StorageGRID . L'utilizzo della CPU per tutti e tre i controller è compreso tra il 75% e il 100%.







SmartStore con controller di archiviazione NetApp : vantaggi per il cliente

- **Disaccoppiamento tra elaborazione e archiviazione.** Splunk SmartStore separa elaborazione e archiviazione, consentendoti di scalarli in modo indipendente.
- **Dati su richiesta.** SmartStore avvicina i dati al calcolo on-demand e fornisce elasticità di elaborazione e archiviazione ed efficienza dei costi per ottenere una conservazione dei dati più lunga su larga scala.
- **Conforme all'API AWS S3.** SmartStore utilizza l'API AWS S3 per comunicare con Restore Storage, che è un archivio di oggetti conforme ad AWS S3 e all'API S3, come StorageGRID.
- **Riduce i requisiti e i costi di archiviazione.** SmartStore riduce i requisiti di archiviazione per i dati obsoleti (caldi/freddi). È necessaria una sola copia dei dati perché lo storage NetApp garantisce la protezione dei dati e gestisce guasti e alta disponibilità.
- **Guasto hardware.** Un errore del nodo in una distribuzione SmartStore non rende i dati inaccessibili e il ripristino dell'indicizzatore in caso di errore hardware o squilibrio dei dati è molto più rapido.
- Cache basata su dati e applicazioni.
- Aggiungi/rimuovi indicizzatori e configura/distruggi cluster su richiesta.

- Il livello di archiviazione non è più legato all'hardware.

Conclusione

Splunk Enterprise è la soluzione SIEM leader di mercato che garantisce risultati concreti nei team di sicurezza, IT e DevOps. L'utilizzo di Splunk è aumentato considerevolmente nelle organizzazioni dei nostri clienti. Pertanto, è necessario aggiungere più fonti di dati, conservando al contempo i dati per un periodo di tempo più lungo, sottoponendo così a stress l'infrastruttura Splunk.

La combinazione di Splunk SmartStore e NetApp StorageGRID è progettata per fornire un'architettura scalabile che consenta alle organizzazioni di ottenere prestazioni di acquisizione migliorate con l'archiviazione di oggetti SmartStore e StorageGRID e una maggiore scalabilità per un ambiente Splunk in più aree geografiche.

Dove trovare ulteriori informazioni

Per saperne di più sulle informazioni descritte nel presente documento, consultare i seguenti documenti e/o siti web:

- "Risorse di documentazione di NetApp StorageGRID"
- "Documentazione del prodotto NetApp"
- "Documentazione di Splunk Enterprise"
- "Splunk Enterprise Informazioni su SmartStore"
- "Manuale di distribuzione distribuita di Splunk Enterprise"
- "Splunk Enterprise Gestione degli indicizzatori e dei cluster di indicizzatori"

Informazioni sul copyright

Informazioni sul marchio commerciale