NetApp StorageGRID avec Splunk SmartStore

NetApp artificial intelligence solutions

NetApp
December 04, 2025

Sommaire

NetApp StorageGRID avec Splunk SmartStore

TR-4869 : NetApp StorageGRID avec Splunk SmartStore

Splunk Enterprise est la solution de gestion des informations et des événements de sécurité (SIEM) leader du marché qui génère des résultats pour les équipes de sécurité, d'informatique et de DevOps.

Aperçu

Les volumes de données continuent de croître à un rythme exponentiel, créant d'énormes opportunités pour les entreprises qui peuvent exploiter cette vaste ressource. Splunk Enterprise continue de gagner en adoption dans une plus grande variété de cas d'utilisation. À mesure que les cas d'utilisation augmentent, la quantité de données que Splunk Enterprise ingère et traite augmente également. L'architecture traditionnelle de Splunk Enterprise est une conception évolutive distribuée offrant un excellent accès aux données et une excellente disponibilité. Cependant, les entreprises utilisant cette architecture sont confrontées à des coûts croissants liés à la mise à l'échelle pour répondre au volume de données en croissance rapide.

Splunk SmartStore avec NetApp StorageGRID résout ce défi en proposant un nouveau modèle de déploiement dans lequel le calcul et le stockage sont découplés. Cette solution offre également une évolutivité et une élasticité inégalées pour les environnements Splunk Enterprise en permettant aux clients de s'adapter à un ou plusieurs sites, tout en réduisant les coûts en permettant au calcul et au stockage de s'adapter indépendamment et en ajoutant une hiérarchisation intelligente au stockage d'objets S3 basé sur le cloud et rentable.

La solution optimise la quantité de données dans le stockage local tout en maintenant les performances de recherche, permettant ainsi de faire évoluer le calcul et le stockage à la demande. SmartStore évalue automatiquement les modèles d'accès aux données pour déterminer quelles données doivent être accessibles pour des analyses en temps réel et quelles données doivent résider dans un stockage d'objets S3 à moindre coût.

Ce rapport technique décrit les avantages que NetApp apporte à une solution Splunk SmartStore tout en démontrant un cadre pour la conception et le dimensionnement de Splunk SmartStore dans votre environnement. Le résultat est une solution simple, évolutive et résiliente qui offre un TCO convaincant. StorageGRID fournit un stockage d'objets basé sur le protocole S3/API évolutif et rentable, également connu sous le nom de stockage à distance, permettant aux organisations de faire évoluer leur solution Splunk à moindre coût tout en augmentant la résilience.

> Splunk SmartStore fait référence au stockage d'objets sous forme de magasins distants ou de niveaux de stockage distants.

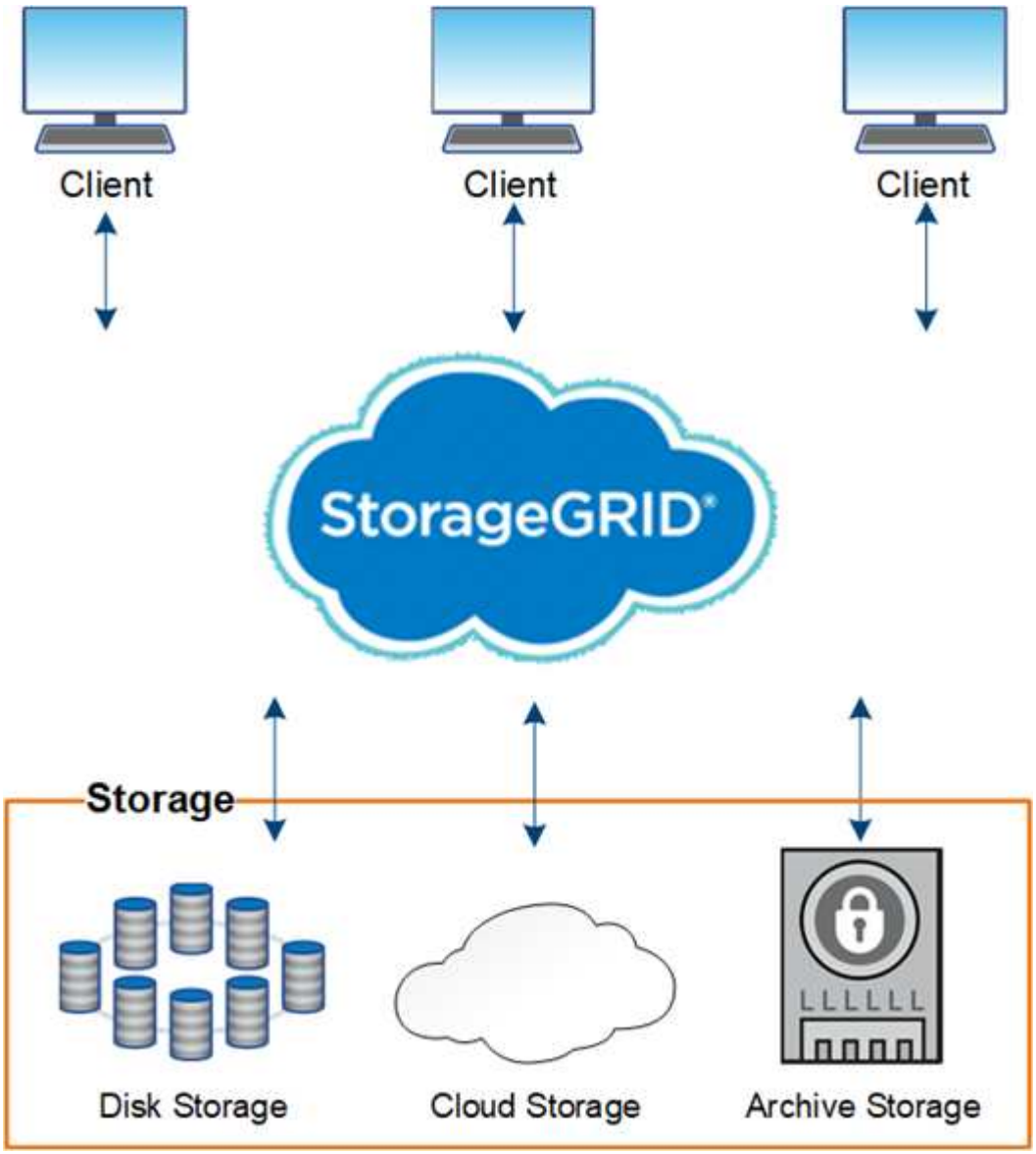
À propos de NetApp StorageGRID

NetApp StorageGRID est une solution de stockage d'objets définie par logiciel pour les grandes archives, les référentiels multimédias et les magasins de données Web. Avec StorageGRID, NetApp s'appuie sur deux décennies d'expérience dans la fourniture de solutions d'innovation et de gestion des données de pointe tout en aidant les organisations à gérer et à maximiser la valeur de leurs informations sur site et dans des déploiements de cloud public, privé ou hybride.

StorageGRID fournit un stockage sécurisé et durable pour les données non structurées à grande échelle. Les politiques de gestion du cycle de vie intégrées et basées sur les métadonnées optimisent l'emplacement de vos données tout au long de leur vie. Le contenu est placé au bon endroit, au bon moment et sur le bon niveau

de stockage pour réduire les coûts. L'espace de noms unique permet d'accéder aux données via un seul appel, quel que soit l'emplacement géographique du stockage StorageGRID . Les clients peuvent déployer et gérer plusieurs instances StorageGRID entre les centres de données et dans l'infrastructure cloud.

Un système StorageGRID est composé de nœuds hétérogènes, redondants et distribués à l'échelle mondiale qui peuvent être intégrés aux applications clientes existantes et de nouvelle génération.



IDC MarketScape a récemment désigné NetApp comme leader dans le dernier rapport, IDC MarketScape : Worldwide Object-Based Storage 2019 Vendor Assessment. Avec près de 20 ans de déploiements de production dans les secteurs les plus exigeants, StorageGRID est un leader reconnu dans le domaine des données non structurées.

Avec StorageGRID, vous pouvez réaliser les objectifs suivants :

- Déployez plusieurs instances StorageGRID pour accéder aux données depuis n'importe quel emplacement entre les centres de données et le cloud via un espace de noms unique qui s'adapte facilement à des centaines de pétaoctets.
- Offrez la flexibilité nécessaire pour déployer et gérer de manière centralisée les infrastructures.
- Offrez une durabilité inégalée avec quinze neuf de durabilité tirant parti du codage d'effacement en

couches (EC).

- Activez davantage de fonctionnalités multicloud hybrides avec des intégrations validées dans Amazon S3 Glacier et Azure Blob.
- Respectez les obligations réglementaires et facilitez la conformité grâce à une conservation des données inviolable, sans API propriétaires ni dépendance vis-à-vis des fournisseurs.

Pour plus d'informations sur la manière dont StorageGRID peut vous aider à résoudre vos problèmes de gestion de données non structurées les plus complexes, consultez le "Page d'accueil de NetApp StorageGRID" .

À propos de Splunk Enterprise

Splunk Enterprise est une plateforme permettant de transformer les données en actions. Les données générées par diverses sources telles que les fichiers journaux, les sites Web, les appareils, les capteurs et les applications sont envoyées et analysées par les indexeurs Splunk, vous permettant de tirer des informations riches des données. Il peut identifier les violations de données, mettre en évidence les tendances des clients et des produits, trouver des opportunités d'optimisation de l'infrastructure ou créer des informations exploitables dans une grande variété de cas d'utilisation.

À propos de Splunk SmartStore

Splunk SmartStore étend les avantages de l'architecture Splunk tout en simplifiant sa capacité à évoluer de manière rentable. Le découplage des ressources de calcul et de stockage donne lieu à des nœuds d'indexation optimisés pour les E/S avec des besoins de stockage considérablement réduits car ils ne stockent qu'un sous-ensemble de données sous forme de cache. Vous n'avez pas besoin d'ajouter de ressources de calcul ou de stockage supplémentaires lorsqu'une seule de ces ressources est nécessaire, ce qui vous permet de réaliser des économies de coûts importantes. Vous pouvez utiliser un stockage d'objets basé sur S3 rentable et facilement évolutif, ce qui simplifie davantage l'environnement, réduit les coûts et vous permet de conserver un ensemble de données plus volumineux.

Splunk SmartStore offre une valeur significative aux organisations, notamment les suivantes :

- Réduire les coûts de stockage en déplaçant les données chaudes vers un stockage d'objets S3 optimisé en termes de coûts
- Mise à l'échelle transparente en découplant le stockage et le calcul
- Simplifier la continuité des activités en exploitant un stockage cloud natif résilient

Présentation de la solution

Cette page décrit les composants utilisés pour compléter cette solution, notamment NetApp StorageGRID, Splunk Enterprise et Splunk SmartStore.

NetApp StorageGRID

NetApp StorageGRID est une plate-forme de stockage d'objets hautes performances et rentable. Il offre une gestion intelligente des données mondiales basée sur des politiques utilisant une architecture de grille distribuée basée sur des nœuds. Il simplifie la gestion de pétaoctets de données non structurées et de milliards d'objets grâce à son espace de noms d'objets global omniprésent combiné à des fonctionnalités de gestion de données sophistiquées. L'accès aux objets par appel unique s'étend sur plusieurs sites et simplifie les architectures à haute disponibilité tout en garantissant un accès continu aux objets, quelles que soient les pannes du site ou de l'infrastructure.

La multilocation permet à plusieurs applications de données non structurées cloud et d'entreprise d'être gérées en toute sécurité au sein de la même grille, augmentant ainsi le retour sur investissement et les cas d'utilisation de StorageGRID. Plusieurs niveaux de service peuvent être créés avec des politiques de cycle de vie d'objets basées sur les métadonnées, optimisant la durabilité, la protection, les performances et la localité dans plusieurs zones géographiques. Les utilisateurs peuvent ajuster les politiques et réaligner le paysage des données de manière non perturbatrice à mesure que leurs besoins évoluent.

SmartStore exploite StorageGRID comme niveau de stockage à distance et permet aux clients de déployer plusieurs sites géographiquement répartis pour une disponibilité et une durabilité robustes, présentées sous la forme d'un espace de noms d'objet unique. Cela permet à Splunk SmartStore de tirer parti des hautes performances, de la capacité dense et de la capacité de StorageGRID à évoluer vers des centaines de nœuds sur plusieurs sites physiques à l'aide d'une seule URL pour interagir avec les objets. Cette URL unique permet également d'étendre le stockage, de mettre à niveau et de réparer sans interruption, même au-delà d'un seul site. Le moteur de politique de gestion des données unique de StorageGRID offre des niveaux optimisés de performances et de durabilité ainsi que le respect des exigences de localisation des données.

Splunk Entreprise

Splunk, leader dans la collecte et l'analyse de données générées par des machines, contribue à simplifier et à moderniser l'informatique grâce à ses capacités d'analyse opérationnelle. Il s'étend également aux cas d'utilisation de l'analyse commerciale, de la sécurité et de l'IoT. Le stockage est un élément essentiel pour un déploiement réussi du logiciel Splunk.

Les données générées par machine constituent le type de big data qui connaît la croissance la plus rapide. Le format est imprévisible et provient de nombreuses sources différentes, souvent à des tarifs élevés et en grands volumes. Ces caractéristiques de charge de travail sont souvent appelées échappement numérique. Splunk SmartStore permet de donner un sens à ces données et fournit une hiérarchisation intelligente des données pour un placement optimisé des données chaudes et tièdes sur le niveau de stockage le plus rentable.

Splunk SmartStore

Splunk SmartStore est une fonctionnalité d'indexation qui utilise le stockage d'objets (également appelé stockage distant ou niveaux de stockage distants) tel que StorageGRID pour stocker des données chaudes à l'aide du protocole S3.

À mesure que le volume de données d'un déploiement augmente, la demande de stockage dépasse généralement la demande de ressources informatiques. SmartStore vous permet de gérer de manière rentable le stockage de votre indexeur et vos ressources de calcul en mettant à l'échelle le calcul et le stockage séparément.

SmartStore introduit un niveau de stockage à distance, utilisant le protocole S3, et un gestionnaire de cache. Ces fonctionnalités permettent aux données de résider localement sur des indexeurs ou sur un stockage distant. Le gestionnaire de cache, qui réside sur l'indexeur, gère le déplacement des données entre l'indexeur et le niveau de stockage distant. Les données sont stockées dans des buckets (chauds et tièdes) avec les métadonnées des buckets.

Avec SmartStore, vous pouvez réduire l'empreinte de stockage de l'indexeur au minimum et choisir des ressources de calcul optimisées pour les E/S, car la plupart des données résident sur le niveau de stockage distant. L'indexeur maintient un cache local, représentant la quantité minimale de données nécessaire pour renvoyer les résultats demandés et prédits. Le cache local contient des buckets chauds, des copies de buckets chauds participant à des recherches actives ou récentes et des métadonnées de bucket.

Splunk SmartStore avec StorageGRID permet aux clients de faire évoluer progressivement l'environnement avec un stockage à distance hautes performances et rentable tout en offrant un degré élevé d'élasticité à la

solution globale. Cela permet aux clients d'ajouter n'importe quel composant (stockage à chaud et/ou stockage S3 chaud) dans n'importe quelle quantité à tout moment, qu'ils aient besoin de plus d'indexeurs, de modifier la conservation des données ou d'augmenter le taux d'ingestion sans aucune interruption.

Avantages de cette solution

La solution permet d'ajouter des ressources de calcul, de stockage à chaud ou S3 pour répondre à la demande croissante en termes de nombre d'utilisateurs ou de taux d'ingestion sur des déploiements mono et multi-sites.

- **Performance.** La combinaison de Splunk SmartStore et de NetApp StorageGRID permet une migration rapide des données entre les buckets chauds et les buckets tièdes à l'aide du stockage d'objets. StorageGRID dynamise le processus de migration en offrant des performances rapides pour les charges de travail d'objets volumineux.
- **Prêt pour le multisite.** L'architecture distribuée StorageGRID permet à Splunk SmartStore d'étendre les déploiements sur des sites uniques et multiples via un espace de noms global unique où les données sont accessibles depuis n'importe quel site, quel que soit l'endroit où se trouvent les données.
- **Évolutivité améliorée.** Faites évoluer les ressources de stockage indépendamment des ressources de calcul pour répondre aux besoins et aux demandes en constante évolution de votre environnement Splunk, offrant ainsi un coût total de possession amélioré.
- **Capacité.** Répondez aux volumes en croissance rapide dans le déploiement de Splunk avec StorageGRID en faisant évoluer un seul espace de noms à plus de 560 Po.
- **Disponibilité des données.** Optimisez la disponibilité des données, les performances, la géodistribution, la conservation, la protection et les coûts de stockage grâce à des politiques basées sur les métadonnées qui peuvent s'ajuster de manière dynamique à mesure que la valeur commerciale de vos données évolue.

Augmentez les performances avec le cache SmartStore, qui est un composant de l'indexeur qui gère le transfert des copies de bucket entre le stockage local (à chaud) et distant (à chaud). Le dimensionnement de Splunk pour cette solution est basé sur le "directives fournies par Splunk" . La solution permet d'ajouter des ressources de calcul, de stockage à chaud ou S3 pour répondre à la demande croissante en termes de nombre d'utilisateurs ou de taux d'ingestion sur des déploiements mono et multi-sites.

Architecture Splunk

Cette section décrit l'architecture Splunk, y compris les définitions clés, les déploiements distribués Splunk, Splunk SmartStore, le flux de données, les exigences matérielles et logicielles, les exigences mono et multisites, etc.

Définitions clés

Les deux tableaux suivants répertorient les composants Splunk et NetApp utilisés dans le déploiement Splunk distribué.

Ce tableau répertorie les composants matériels Splunk pour la configuration distribuée de Splunk Enterprise.

Composant Splunk	Tâche
Indexeur	Référentiel pour les données Splunk Enterprise

Composant Splunk	Tâche
Transitaire universel	Responsable de l'ingestion des données et de leur transmission aux indexeurs
Tête de recherche	L'interface utilisateur utilisée pour rechercher des données dans les indexeurs
Maître de cluster	Gère l'installation Splunk des indexeurs et des têtes de recherche
Console de surveillance	Outil de surveillance centralisé utilisé sur l'ensemble du déploiement
Licence master	Le maître des licences gère les licences Splunk Enterprise
Serveur de déploiement	Met à jour les configurations et distribue les applications au composant de traitement
Composant de stockage	Tâche
NetApp AFF	Stockage entièrement flash utilisé pour gérer les données de niveau chaud. Également connu sous le nom de stockage local.
NetApp StorageGRID	Stockage d'objets S3 utilisé pour gérer les données de niveau chaud. Utilisé par SmartStore pour déplacer des données entre les niveaux chaud et tiède. Également connu sous le nom de stockage à distance.

Ce tableau répertorie les composants de l'architecture de stockage Splunk.

Composant Splunk	Tâche	Composant responsable
Magasin intelligent	Fournit aux indexeurs la possibilité de hiérarchiser les données du stockage local vers le stockage d'objets.	Splunk
Chaud	Le point d'atterrissage où les transitaires universels placent les données nouvellement écrites. Le stockage est accessible en écriture et les données sont consultables. Ce niveau de données est généralement composé de SSD ou de disques durs rapides.	ONTAP
Gestionnaire de cache	Gère le cache local des données indexées, récupère les données chaudes du stockage distant lorsqu'une recherche se produit et supprime les données les moins fréquemment utilisées du cache.	Magasin intelligent

Composant Splunk	Tâche	Composant responsable
Chaud	Les données sont transférées logiquement vers le bucket, renommées d'abord vers le niveau chaud à partir du niveau chaud. Les données de ce niveau sont protégées et, comme le niveau chaud, peuvent être composées de SSD ou de disques durs de plus grande capacité. Les sauvegardes incrémentielles et complètes sont prises en charge à l'aide de solutions de protection des données courantes.	StorageGRID

Déploiements distribués Splunk

Pour prendre en charge des environnements plus vastes dans lesquels les données proviennent de nombreuses machines, vous devez traiter de grands volumes de données. Si de nombreux utilisateurs doivent rechercher les données, vous pouvez faire évoluer le déploiement en distribuant les instances Splunk Enterprise sur plusieurs machines. C'est ce qu'on appelle un déploiement distribué.

Dans un déploiement distribué typique, chaque instance Splunk Enterprise exécute une tâche spécialisée et réside sur l'un des trois niveaux de traitement correspondant aux principales fonctions de traitement.

Le tableau suivant répertorie les niveaux de traitement de Splunk Enterprise.

Étage	Composant	Description
Saisie de données	Transitaire	Un transitaire consomme des données, puis les transmet à un groupe d'indexeurs.
Indexage	Indexeur	Un indexeur indexe les données entrantes qu'il reçoit généralement d'un groupe de transitaires. L'indexeur transforme les données en événements et stocke les événements dans un index. L'indexeur recherche également les données indexées en réponse aux demandes de recherche d'une tête de recherche.
Gestion de la recherche	Tête de recherche	Une tête de recherche sert de ressource centrale pour la recherche. Les têtes de recherche d'un cluster sont interchangeables et ont accès aux mêmes recherches, tableaux de bord, objets de connaissances, etc., à partir de n'importe quel membre du cluster de têtes de recherche.

Le tableau suivant répertorie les composants importants utilisés dans un environnement Splunk Enterprise distribué.

Composant	Description	Responsabilité
Maître du cluster d'index	Coordonne les activités et les mises à jour d'un cluster d'indexeurs	Gestion des indices
cluster d'index	Groupe d'indexeurs Splunk Enterprise configurés pour répliquer des données entre eux	Indexage
Déploiement de la tête de recherche	Gère le déploiement et les mises à jour du cluster maître	Gestion de la tête de recherche
Cluster de têtes de recherche	Groupe de têtes de recherche qui sert de ressource centrale pour la recherche	Gestion de la recherche
Équilibreurs de charge	Utilisé par les composants en cluster pour gérer la demande croissante des têtes de recherche, des indexeurs et de la cible S3 afin de répartir la charge sur les composants en cluster.	Gestion de la charge pour les composants en cluster

Découvrez les avantages suivants des déploiements distribués Splunk Enterprise :

- Accéder à des sources de données diverses ou dispersées
- Fournir des fonctionnalités pour gérer les besoins en données des entreprises de toute taille et de toute complexité
- Obtenez une haute disponibilité et assurez la reprise après sinistre grâce à la réplication des données et au déploiement multisite

Splunk SmartStore

SmartStore est une fonctionnalité d'indexation qui permet aux magasins d'objets distants tels qu'Amazon S3 de stocker des données indexées. À mesure que le volume de données d'un déploiement augmente, la demande de stockage dépasse généralement la demande de ressources de calcul. SmartStore vous permet de gérer de manière rentable le stockage de votre indexeur et vos ressources de calcul en mettant à l'échelle ces ressources séparément.

SmartStore introduit un niveau de stockage à distance et un gestionnaire de cache. Ces fonctionnalités permettent aux données de résider soit localement sur des indexeurs, soit sur le niveau de stockage distant. Le gestionnaire de cache gère le déplacement des données entre l'indexeur et le niveau de stockage distant, qui est configuré sur l'indexeur.

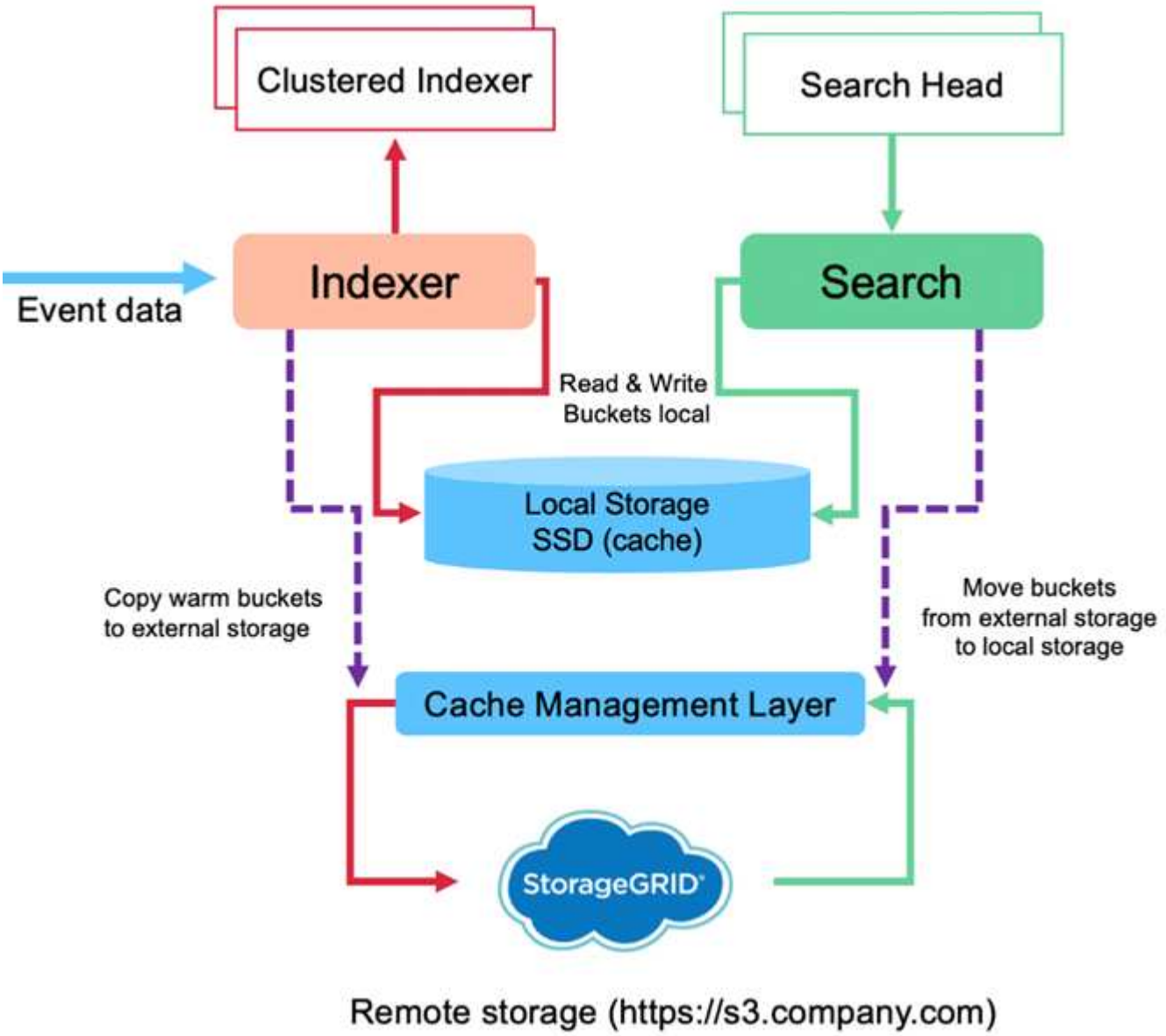
Avec SmartStore, vous pouvez réduire au minimum l'empreinte de stockage de l'indexeur et choisir des ressources de calcul optimisées pour les E/S. La plupart des données résident sur le stockage distant. L'indexeur conserve un cache local contenant une quantité minimale de données : buckets chauds, copies de buckets chauds participant à des recherches actives ou récentes et métadonnées de bucket.

Flux de données Splunk SmartStore

Lorsque les données provenant de diverses sources atteignent les indexeurs, les données sont indexées et enregistrées localement dans un bucket chaud. L'indexeur réplique également les données du compartiment chaud vers les indexeurs cibles. Jusqu'à présent, le flux de données est identique au flux de données des index non SmartStore.

Lorsque le seau chaud devient chaud, le flux de données diverge. L'indexeur source copie le bucket chaud dans le magasin d'objets distant (niveau de stockage distant) tout en laissant la copie existante dans son cache, car les recherches ont tendance à s'exécuter sur des données récemment indexées. Cependant, les indexeurs cibles suppriment leurs copies car le magasin distant offre une haute disponibilité sans conserver plusieurs copies locales. La copie principale du bucket réside désormais dans le magasin distant.

L'image suivante montre le flux de données Splunk SmartStore.



Le gestionnaire de cache sur l'indexeur est au cœur du flux de données SmartStore. Il récupère des copies des buckets du magasin distant si nécessaire pour gérer les demandes de recherche. Il supprime également

les copies plus anciennes ou moins recherchées des buckets du cache, car la probabilité qu'ils participent aux recherches diminue avec le temps.

Le travail du gestionnaire de cache est d'optimiser l'utilisation du cache disponible tout en garantissant que les recherches ont un accès immédiat aux compartiments dont elles ont besoin.

Configuration logicielle requise

Le tableau ci-dessous répertorie les composants logiciels nécessaires à la mise en œuvre de la solution. Les composants logiciels utilisés dans toute implémentation de la solution peuvent varier en fonction des exigences du client.

Famille de produits	Nom du produit	Version du produit	Système opérateur
NetApp StorageGRID	Stockage d'objets StorageGRID	11,6	n / A
CentOS	CentOS	8,1	CentOS 7.x
Splunk Entreprise	Splunk Enterprise avec SmartStore	8.0.3	CentOS 7.x

Exigences mono et multisites

Dans un environnement Splunk Enterprise (déploiements moyens et grands) où les données proviennent de nombreuses machines et où de nombreux utilisateurs doivent rechercher les données, vous pouvez faire évoluer votre déploiement en distribuant des instances Splunk Enterprise sur un ou plusieurs sites.

Découvrez les avantages suivants des déploiements distribués Splunk Enterprise :

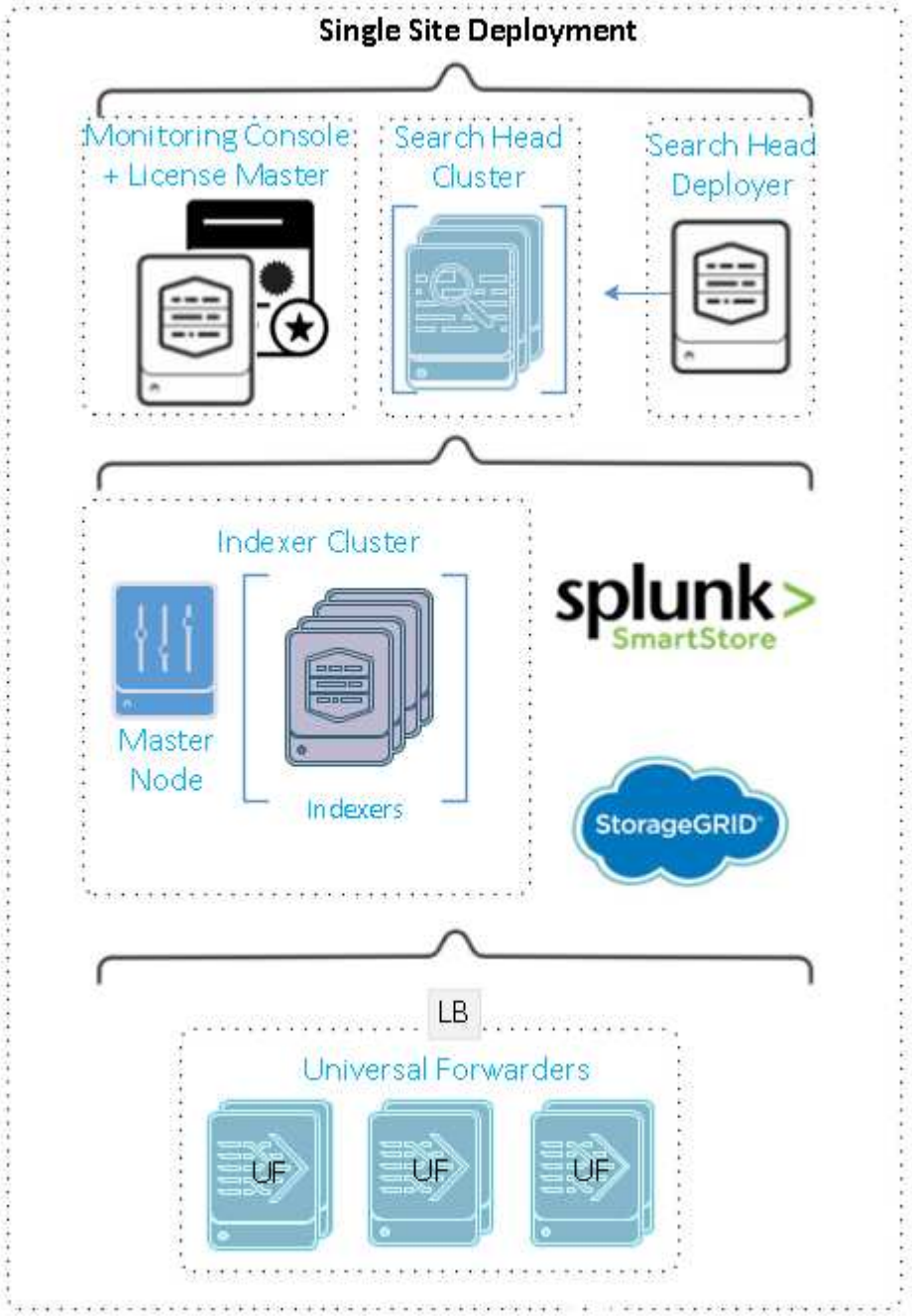
- Accéder à des sources de données diverses ou dispersées
- Fournir des fonctionnalités pour gérer les besoins en données des entreprises de toute taille et de toute complexité
- Obtenez une haute disponibilité et assurez la reprise après sinistre grâce à la réplication des données et au déploiement multisite

Le tableau suivant répertorie les composants utilisés dans un environnement Splunk Enterprise distribué.

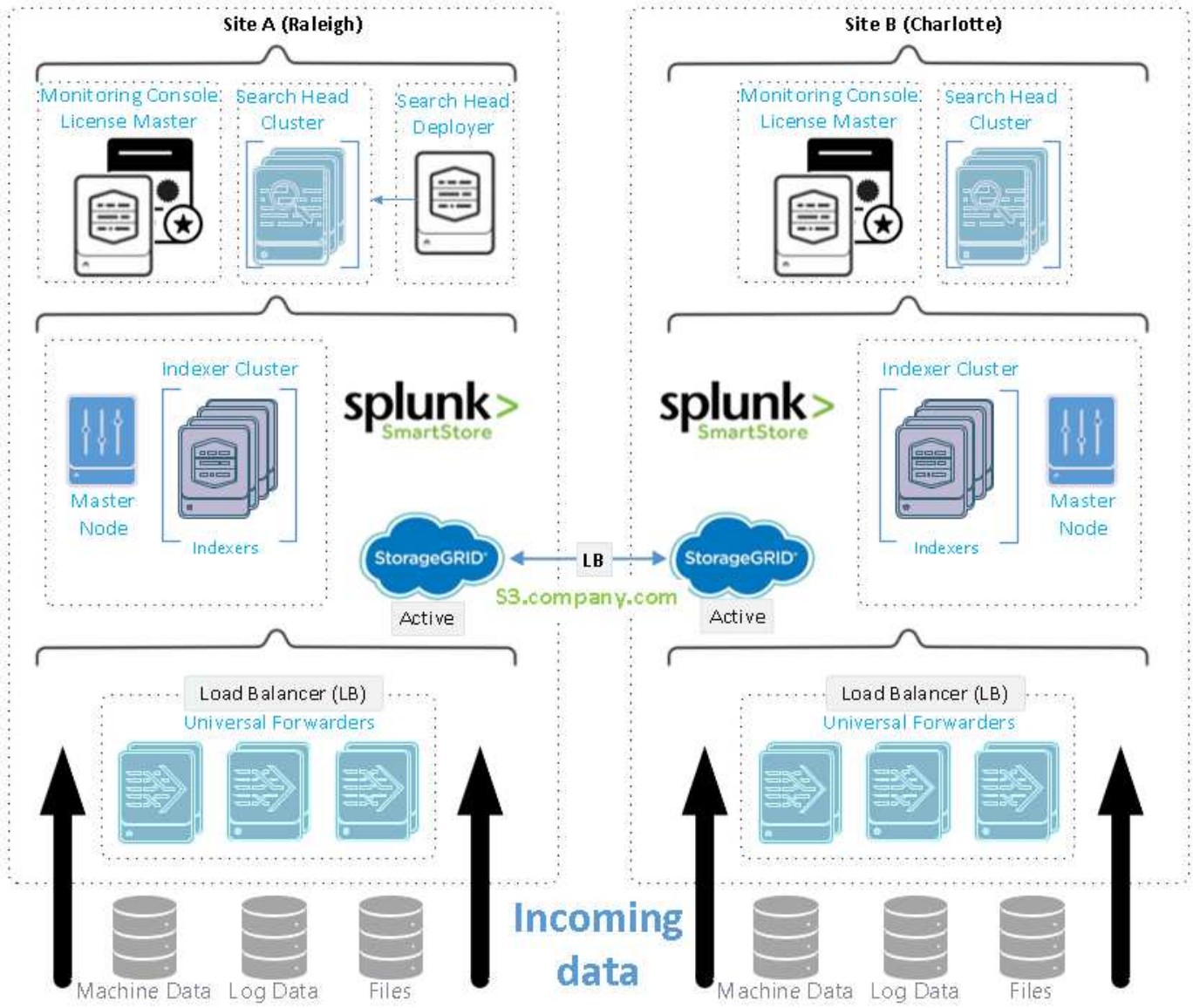
Composant	Description	Responsabilité
Maître du cluster d'index	Coordonne les activités et les mises à jour d'un cluster d'indexeurs	Gestion des indices
cluster d'index	Groupe d'indexeurs Splunk Enterprise configurés pour répliquer les données des autres	Indexage
Déploiement de la tête de recherche	Gère le déploiement et les mises à jour du cluster maître	Gestion de la tête de recherche
Cluster de têtes de recherche	Groupe de têtes de recherche qui sert de ressource centrale pour la recherche	Gestion de la recherche

Composant	Description	Responsabilité
Équilibreurs de charge	Utilisé par les composants en cluster pour gérer la demande croissante des têtes de recherche, des indexeurs et de la cible S3 afin de répartir la charge sur les composants en cluster.	Gestion de la charge pour les composants en cluster

Cette figure illustre un exemple de déploiement distribué sur un seul site.



Cette figure illustre un exemple de déploiement distribué multisite.



Configuration matérielle requise

Les tableaux suivants répertorient le nombre minimum de composants matériels requis pour implémenter la solution. Les composants matériels utilisés dans les implémentations spécifiques de la solution peuvent varier en fonction des exigences du client.

> Que vous ayez déployé Splunk SmartStore et StorageGRID sur un seul site ou sur plusieurs sites, tous les systèmes sont gérés à partir de StorageGRID GRID Manager dans une seule fenêtre. Consultez la section « Gestion simple avec Grid Manager » pour plus de détails.

Ce tableau répertorie le matériel utilisé pour un seul site.

Matériel	Quantité	Disque	Capacité utilisable	Remarque
StorageGRID SG1000	1	n / A	n / A	Nœud d'administration et équilibreur de charge

Matériel	Quantité	Disque	Capacité utilisable	Remarque
StorageGRID SG6060	4	x48, 8 To (disque dur NL-SAS)	1PB	Stockage à distance

Ce tableau répertorie le matériel utilisé pour une configuration multisite (par site).

Matériel	Quantité	Disque	Capacité utilisable	Remarque
StorageGRID SG1000	2	n / A	n / A	Nœud d'administration et équilibreur de charge
StorageGRID SG6060	4	x48, 8 To (disque dur NL-SAS)	1PB	Stockage à distance

Équilibreur de charge NetApp StorageGRID : SG1000

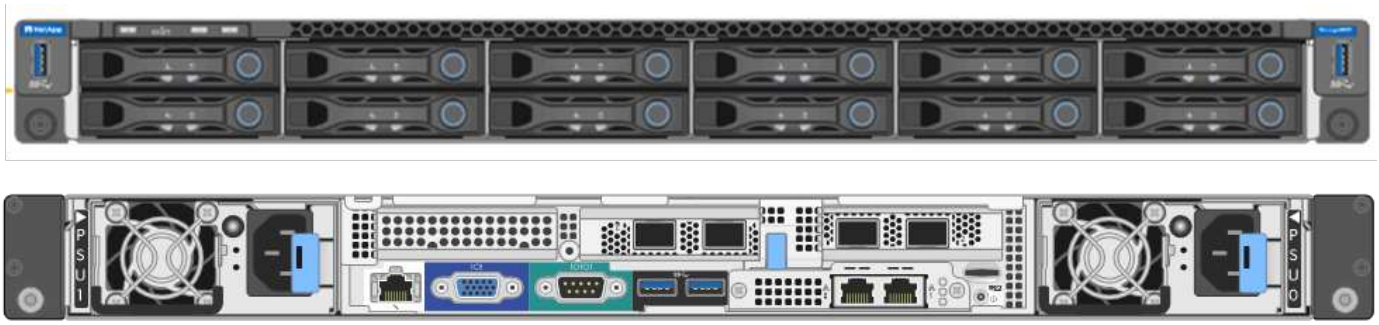
Le stockage d'objets nécessite l'utilisation d'un équilibreur de charge pour présenter l'espace de noms de stockage cloud. StorageGRID prend en charge les équilibreurs de charge tiers des principaux fournisseurs tels que F5 et Citrix, mais de nombreux clients choisissent l'équilibreur StorageGRID de niveau entreprise pour sa simplicité, sa résilience et ses hautes performances. L'équilibreur de charge StorageGRID est disponible sous forme de machine virtuelle, de conteneur ou d'appliance spécialement conçue.

Le StorageGRID SG1000 facilite l'utilisation de groupes de haute disponibilité (HA) et l'équilibrage de charge intelligent pour les connexions de chemin de données S3. Aucun autre système de stockage d'objets sur site ne fournit un équilibreur de charge personnalisé.

L'appareil SG1000 offre les fonctionnalités suivantes :

- Un équilibreur de charge et, éventuellement, des fonctions de nœud d'administration pour un système StorageGRID
- Le programme d'installation de l'appliance StorageGRID pour simplifier le déploiement et la configuration des nœuds
- Configuration simplifiée des points de terminaison S3 et SSL
- Bande passante dédiée (par rapport au partage d'un équilibreur de charge tiers avec d'autres applications)
- Jusqu'à 4 x 100 Gbit/s de bande passante Ethernet agrégée

L'image suivante montre l'appareil SG1000 Gateway Services.

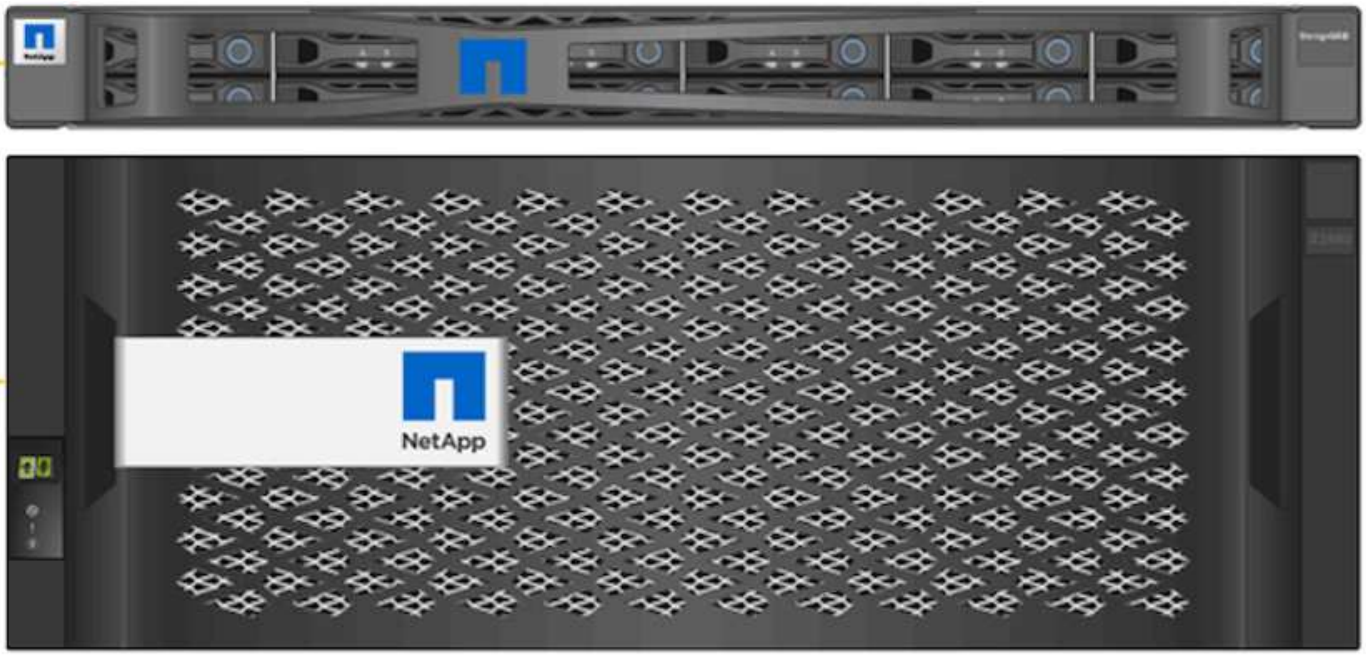


SG6060

L'appliance StorageGRID SG6060 comprend un contrôleur de calcul (SG6060) et une étagère de contrôleur de stockage (E-Series E2860) contenant deux contrôleurs de stockage et 60 disques. Cet appareil offre les fonctionnalités suivantes :

- Évoluez jusqu'à 400 Po dans un seul espace de noms.
- Jusqu'à 4x 25 Gbit/s de bande passante Ethernet agrégée.
- Inclut le programme d'installation de l'appliance StorageGRID pour simplifier le déploiement et la configuration des nœuds.
- Chaque appareil SG6060 peut disposer d'une ou deux étagères d'extension supplémentaires pour un total de 180 disques.
- Deux contrôleurs E-Series E2800 (configuration duplex) pour fournir une prise en charge du basculement du contrôleur de stockage.
- Étagère à cinq tiroirs pouvant contenir soixante disques de 3,5 pouces (deux disques SSD et 58 disques NL-SAS).

L'image suivante montre l'appareil SG6060.



Conception de Splunk

Le tableau suivant répertorie la configuration Splunk pour un seul site.

Composant Splunk	Tâche	Quantité	Noyaux	Mémoire	Système d'exploitation
Transitaire universel	Responsable de l'ingestion des données et de leur transmission aux indexeurs	4	16 cœurs	32 Go de RAM	CentOS 8.1

Composant Splunk	Tâche	Quantité	Noyaux	Mémoire	Système d'exploitation
Indexeur	Gère les données des utilisateurs	10	16 cœurs	32 Go de RAM	CentOS 8.1
Tête de recherche	L'interface utilisateur recherche des données dans les indexeurs	3	16 cœurs	32 Go de RAM	CentOS 8.1
Déploiement de la tête de recherche	Gère les mises à jour des clusters de têtes de recherche	1	16 cœurs	32 Go de RAM	CentOS 8.1
Maître de cluster	Gère l'installation et les indexeurs de Splunk	1	16 cœurs	32 Go de RAM	CentOS 8.1
Console de surveillance et maître de licence	Effectue une surveillance centralisée de l'ensemble du déploiement Splunk et gère les licences Splunk	1	16 cœurs	32 Go de RAM	CentOS 8.1

Les tableaux suivants décrivent la configuration Splunk pour les configurations multisites.

Ce tableau répertorie la configuration Splunk pour une configuration multisite (site A).

Composant Splunk	Tâche	Quantité	Noyaux	Mémoire	Système d'exploitation
Transitaire universel	Responsable de l'ingestion des données et de leur transmission aux indexeurs.	4	16 cœurs	32 Go de RAM	CentOS 8.1
Indexeur	Gère les données des utilisateurs	10	16 cœurs	32 Go de RAM	CentOS 8.1
Tête de recherche	L'interface utilisateur recherche des données dans les indexeurs	3	16 cœurs	32 Go de RAM	CentOS 8.1

Composant Splunk	Tâche	Quantité	Noyaux	Mémoire	Système d'exploitation
Déploiement de la tête de recherche	Gère les mises à jour des clusters de têtes de recherche	1	16 cœurs	32 Go de RAM	CentOS 8.1
Maître de cluster	Gère l'installation et les indexeurs de Splunk	1	16 cœurs	32 Go de RAM	CentOS 8.1
Console de surveillance et maître de licence	Effectue une surveillance centralisée de l'ensemble du déploiement Splunk et gère les licences Splunk.	1	16 cœurs	32 Go de RAM	CentOS 8.1

Ce tableau répertorie la configuration Splunk pour une configuration multisite (site B).

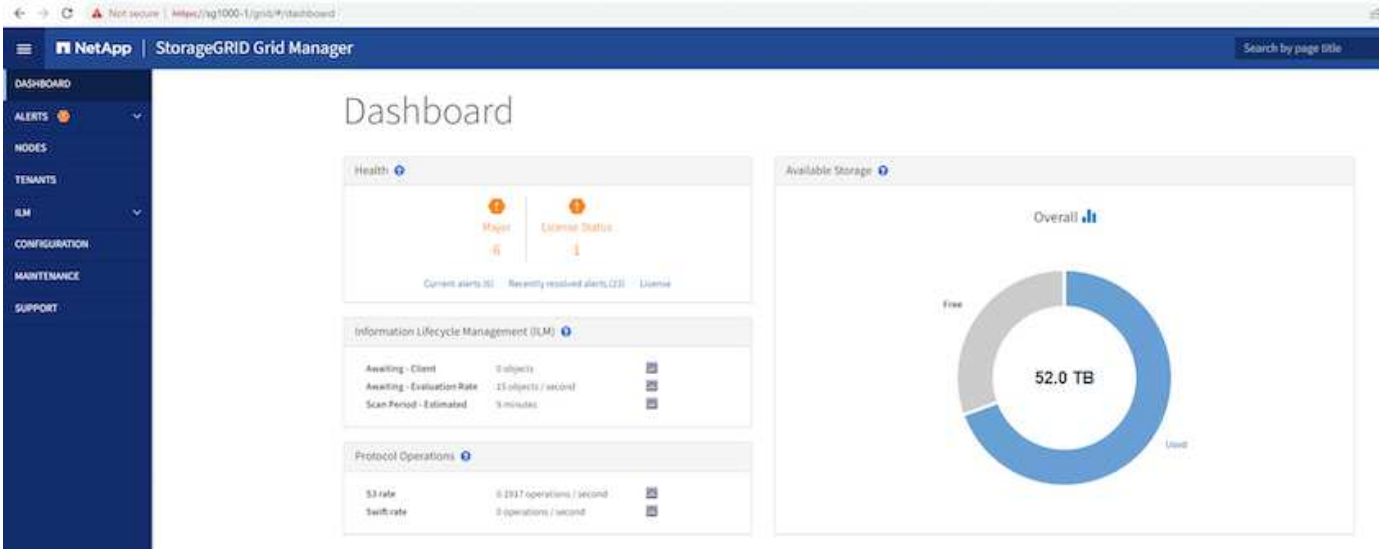
Composant Splunk	Tâche	Quantité	Noyaux	Mémoire	Système d'exploitation
Transitaire universel	Responsable de l'ingestion des données et de leur transmission aux indexeurs	4	16 cœurs	32 Go de RAM	CentOS 8.1
Indexeur	Gère les données des utilisateurs	10	16 cœurs	32 Go de RAM	CentOS 8.1
Tête de recherche	L'interface utilisateur recherche des données dans les indexeurs	3	16 cœurs	32 Go de RAM	CentOS 8.1
Maître de cluster	Gère l'installation et les indexeurs de Splunk	1	16 cœurs	32 Go de RAM	CentOS 8.1
Console de surveillance et maître de licence	Effectue une surveillance centralisée de l'ensemble du déploiement Splunk et gère les licences Splunk	1	16 cœurs	32 Go de RAM	CentOS 8.1

Fonctionnalités flexibles de StorageGRID pour Splunk SmartStore

StorageGRID dispose d'une grande variété de fonctionnalités que les utilisateurs peuvent exploiter et personnaliser pour leur environnement en constante évolution. Du déploiement à la mise à l'échelle de votre Splunk SmartStore, votre environnement exige une adoption rapide des changements et ne doit pas perturber Splunk. Les politiques de gestion des données flexibles (ILM) et les classificateurs de trafic (QoS) de StorageGRID vous permettent de planifier et de vous adapter à votre environnement.

Gestion simple avec Grid Manager

Grid Manager est l'interface graphique basée sur un navigateur qui vous permet de configurer, de gérer et de surveiller votre système StorageGRID sur des emplacements distribués à l'échelle mondiale dans un seul volet de verre, comme illustré dans l'image suivante.



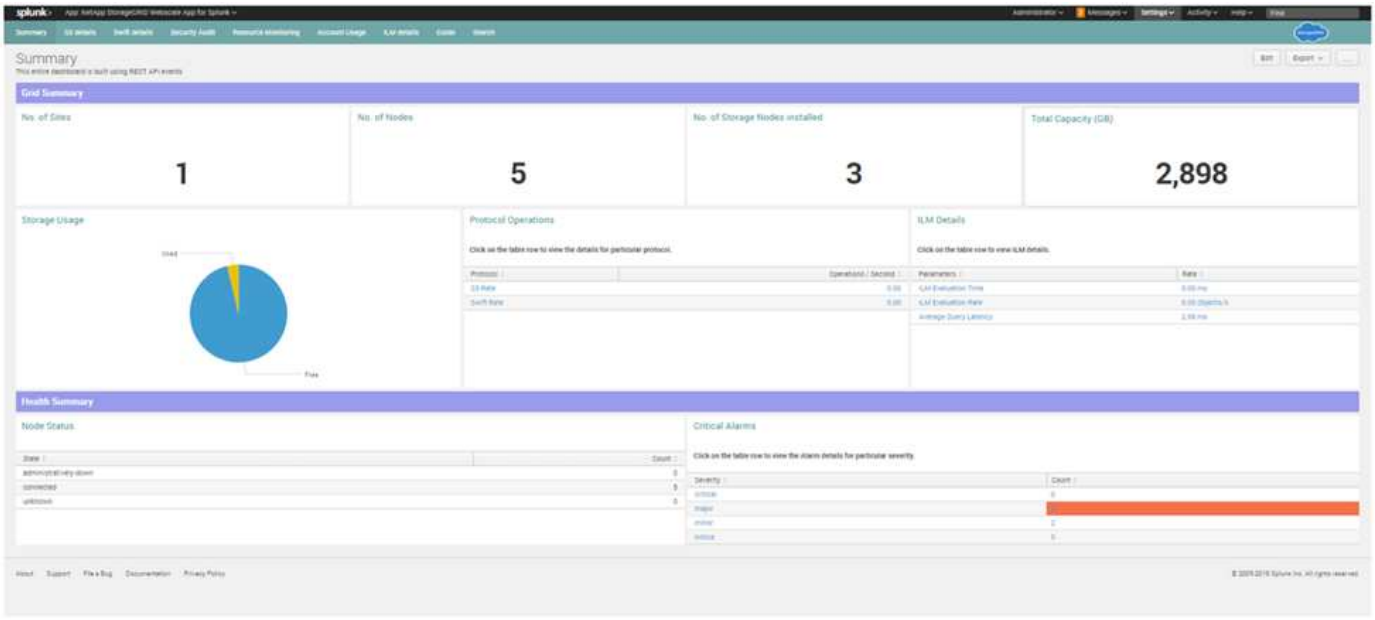
Effectuez les tâches suivantes avec l'interface Grid Manager :

- Gérez des référentiels d'objets distribués à l'échelle mondiale et à l'échelle du pétaoctet, tels que des images, des vidéos et des enregistrements.
- Surveillez les nœuds et les services de la grille pour garantir la disponibilité des objets.
- Gérez le placement des données d'objet au fil du temps à l'aide de règles de gestion du cycle de vie des informations (ILM). Ces règles régissent ce qui arrive aux données d'un objet après son ingestion, comment elles sont protégées contre la perte, où les données de l'objet sont stockées et pendant combien de temps.
- Surveiller les transactions, les performances et les opérations au sein du système.

Application NetApp StorageGRID pour Splunk

L'application NetApp StorageGRID pour Splunk est une application spécifique à Splunk Enterprise. Cette application fonctionne en conjonction avec le module complémentaire NetApp StorageGRID pour Splunk. Il offre une visibilité sur l'état de StorageGRID , les informations d'utilisation du compte, les détails de l'audit de sécurité, l'utilisation et la surveillance des ressources, etc.

L’image suivante montre l’application StorageGRID pour Splunk.



Politiques ILM

StorageGRID dispose de politiques de gestion des données flexibles qui incluent la conservation de plusieurs copies de vos objets et l’utilisation de schémas EC (codage d’effacement) tels que 2+1 et 4+2 (et bien d’autres) pour stocker vos objets en fonction des exigences spécifiques de performances et de protection des données. À mesure que les charges de travail et les exigences évoluent au fil du temps, il est courant que les politiques ILM doivent également évoluer au fil du temps. La modification des politiques ILM est une fonctionnalité essentielle, permettant aux clients de StorageGRID de s’adapter rapidement et facilement à leur environnement en constante évolution.

Performances

StorageGRID adapte les performances en ajoutant davantage de nœuds, qui peuvent être des machines virtuelles, des appareils bare metal ou des appliances spécialement conçues comme les SG5712, SG5760, SG6060 ou SGF6024. Lors de nos tests, nous avons dépassé les exigences de performances clés de SmartStore avec une grille à trois nœuds de taille minimale en utilisant l’appliance SG6060. À mesure que les clients font évoluer leur infrastructure Splunk avec des indexeurs supplémentaires, ils peuvent ajouter davantage de nœuds de stockage pour augmenter les performances et la capacité.

Configuration de l’équilibreur de charge et du point de terminaison

Les nœuds d’administration de StorageGRID fournissent l’interface utilisateur de Grid Manager et le point de terminaison de l’API REST pour afficher, configurer et gérer votre système StorageGRID , ainsi que des journaux d’audit pour suivre l’activité du système. Pour fournir un point de terminaison S3 hautement disponible pour le stockage à distance Splunk SmartStore, nous avons implémenté l’équilibreur de charge StorageGRID , qui s’exécute en tant que service sur les nœuds d’administration et les nœuds de passerelle. De plus, l’équilibreur de charge gère également le trafic local et communique avec le GSLB (Global Server Load Balancing) pour faciliter la reprise après sinistre.

Pour améliorer davantage la configuration des points de terminaison, StorageGRID fournit des stratégies de classification du trafic intégrées au nœud d’administration, vous permet de surveiller le trafic de votre charge de travail et d’appliquer diverses limites de qualité de service (QoS) à vos charges de travail. Les stratégies de classification du trafic sont appliquées aux points de terminaison sur le service StorageGRID Load Balancer

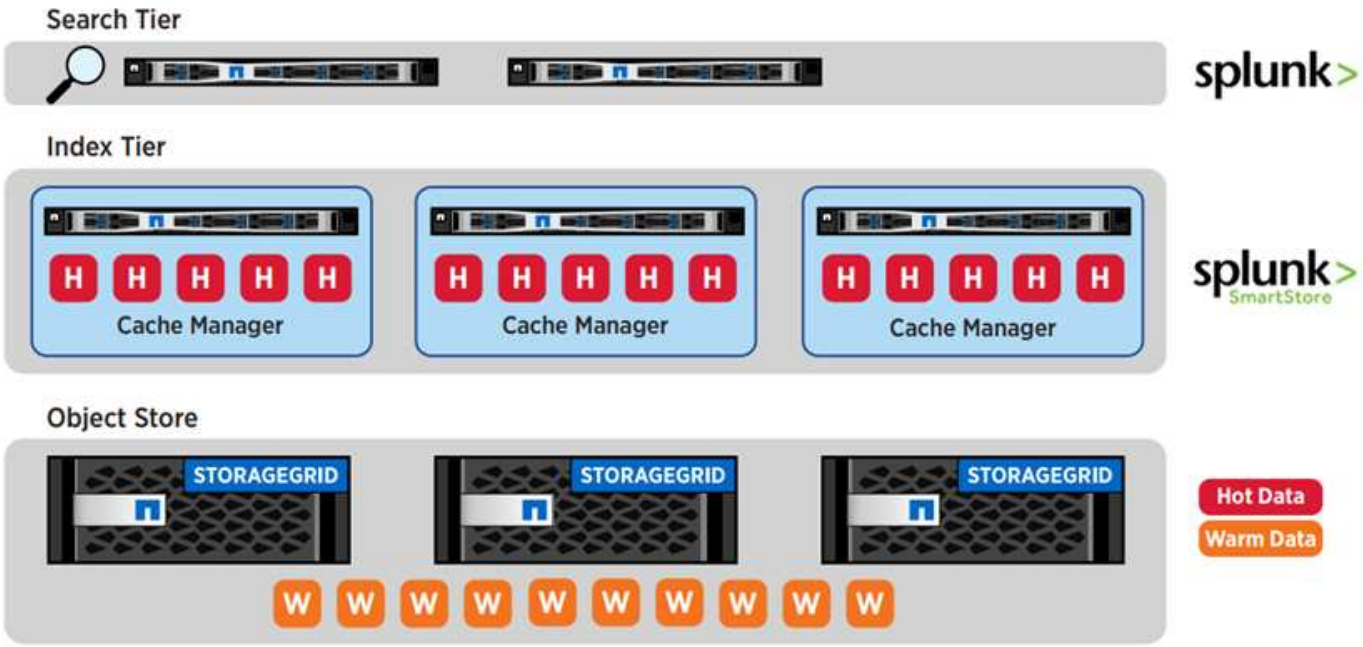
pour les nœuds de passerelle et les nœuds d'administration. Ces politiques peuvent aider à limiter et à surveiller le trafic.

Hiérarchisation intelligente et économies de coûts

À mesure que les clients réalisent la puissance et la facilité d'utilisation de l'analyse de données Splunk, ils souhaitent naturellement indexer une quantité toujours croissante de données. À mesure que la quantité de données augmente, l'infrastructure de calcul et de stockage nécessaire pour les gérer augmente également. Étant donné que les données plus anciennes sont référencées moins fréquemment, engager la même quantité de ressources de calcul et consommer un stockage primaire coûteux devient de plus en plus inefficace. Pour fonctionner à grande échelle, les clients bénéficient du déplacement des données chaudes vers un niveau plus rentable, libérant ainsi du calcul et du stockage principal pour les données chaudes.

Splunk SmartStore avec StorageGRID offre aux organisations une solution évolutive, performante et rentable. Étant donné que SmartStore est sensible aux données, il évalue automatiquement les modèles d'accès aux données pour déterminer quelles données doivent être accessibles pour l'analyse en temps réel (données chaudes) et quelles données doivent résider dans un stockage à long terme à moindre coût (données chaudes). SmartStore utilise l'API AWS S3 standard de l'industrie de manière dynamique et intelligente, en plaçant les données dans le stockage S3 fourni par StorageGRID. L'architecture évolutive flexible de StorageGRID permet au niveau de données chaudes de croître de manière rentable selon les besoins. L'architecture basée sur les nœuds de StorageGRID garantit que les exigences de performances et de coûts sont satisfaites de manière optimale.

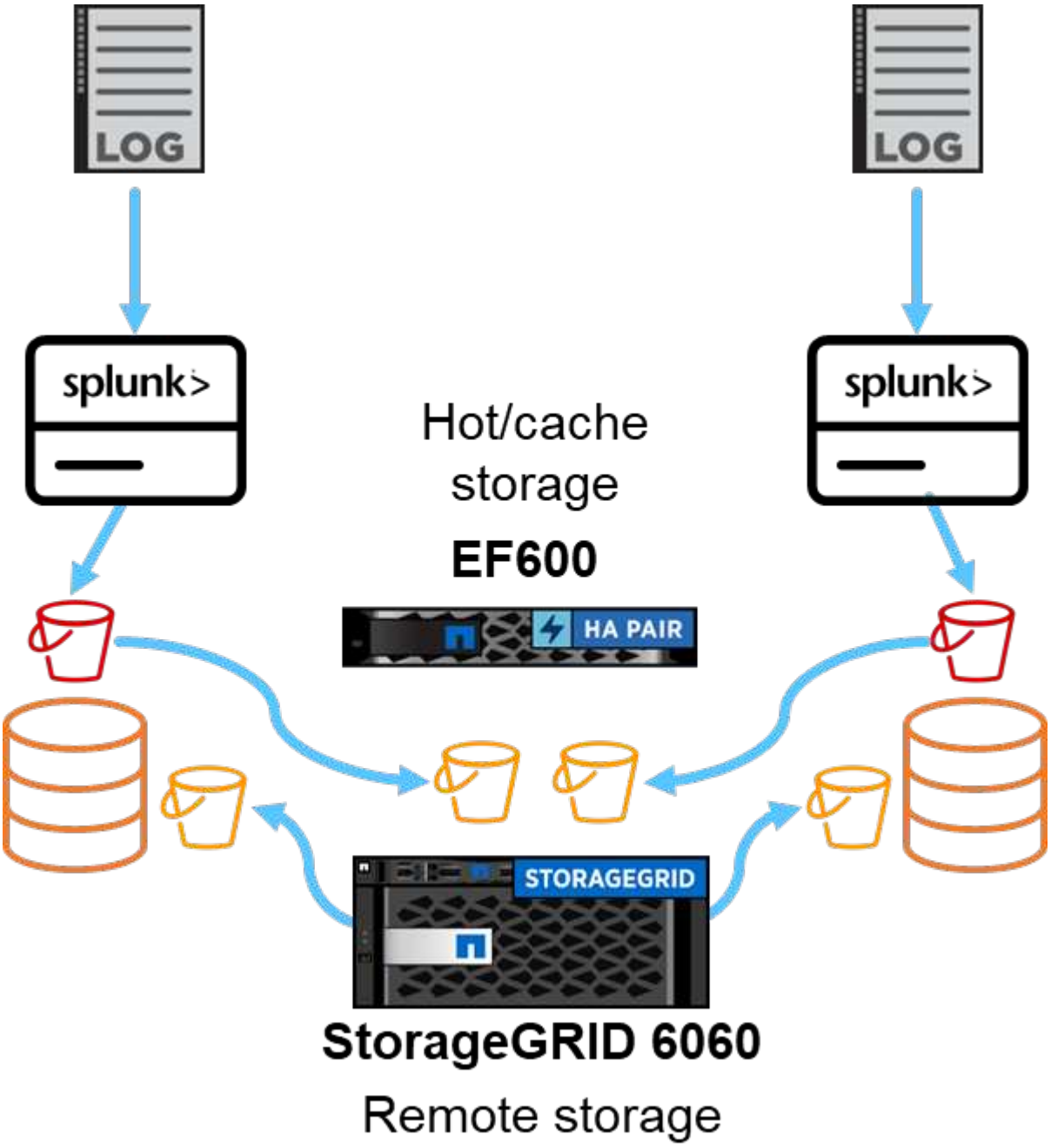
L'image suivante illustre la hiérarchisation de Splunk et StorageGRID .



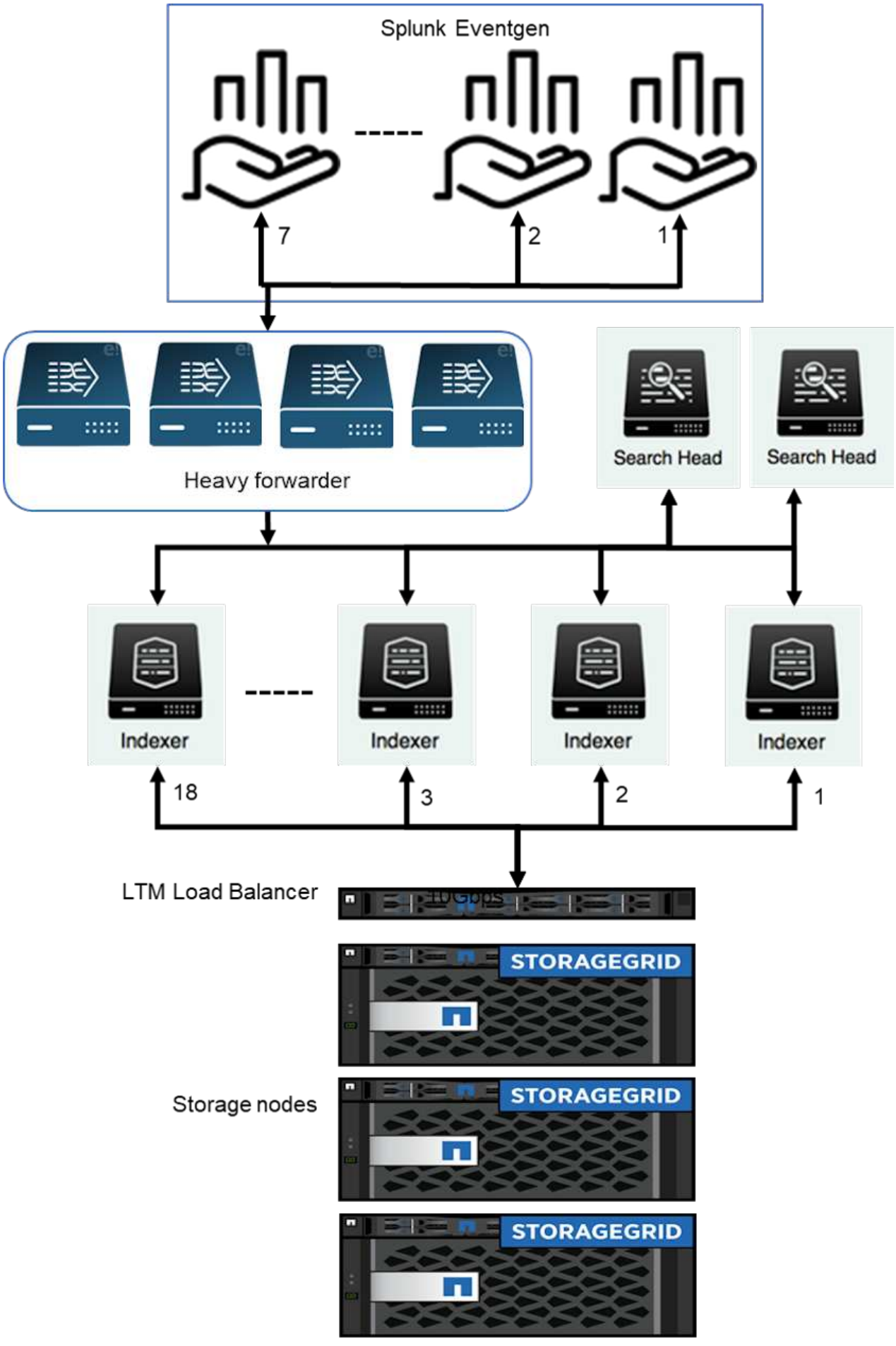
La combinaison leader du secteur de Splunk SmartStore avec NetApp StorageGRID offre les avantages d'une architecture découplée via une solution complète.

Performances du SmartStore sur un seul site

Cette section décrit les performances de Splunk SmartStore sur un contrôleur NetApp StorageGRID . Splunk SmartStore déplace les données chaudes vers un stockage distant, qui dans ce cas est le stockage d'objets StorageGRID dans la validation des performances.



Nous avons utilisé EF600 pour le stockage à chaud/cache et StorageGRID 6060 pour le stockage à distance. Nous avons utilisé l'architecture suivante pour la validation des performances. Nous avons utilisé deux têtes de recherche, quatre transitaires lourds pour transmettre les données aux indexeurs, sept générateurs d'événements Splunk (Eventgens) pour générer les données en temps réel et 18 indexeurs pour stocker les données.


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

Configuration

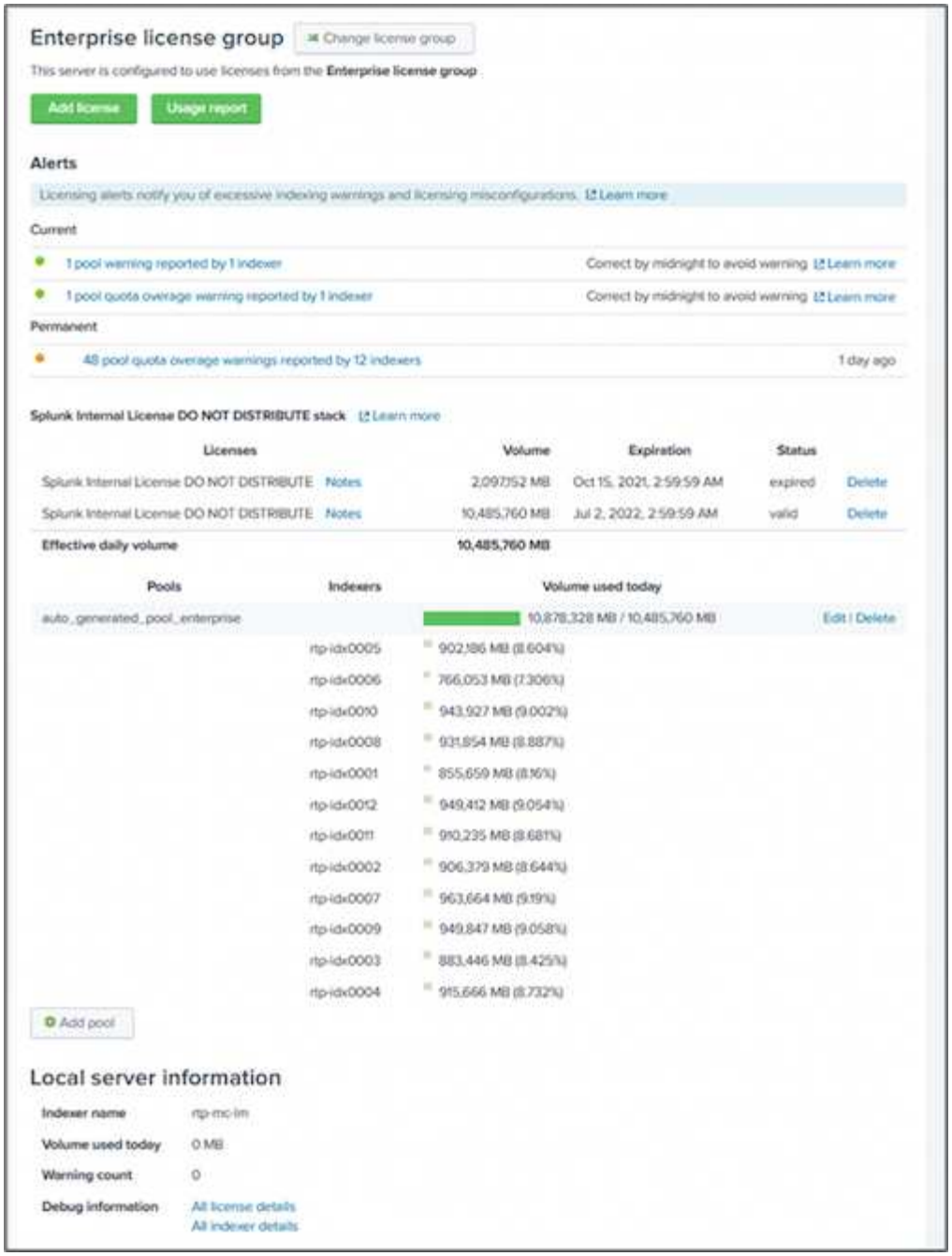
Ce tableau répertorie le matériel utilisé pour la validation des performances de SmartStorage.

Composant Splunk	Tâche	Quantité	Noyaux	Mémoire	Système d'exploitation
Porteur lourd	Responsable de l'ingestion des données et de leur transmission aux indexeurs	4	16 cœurs	32 Go de RAM	SLED 15 SP2
Indexeur	Gère les données des utilisateurs	18	16 cœurs	32 Go de RAM	SLED 15 SP2
Tête de recherche	L'interface utilisateur recherche des données dans les indexeurs	2	16 cœurs	32 Go de RAM	SLED 15 SP2
Déploiement de la tête de recherche	Gère les mises à jour des clusters de têtes de recherche	1	16 cœurs	32 Go de RAM	SLED 15 SP2
Maître de cluster	Gère l'installation et les indexeurs de Splunk	1	16 cœurs	32 Go de RAM	SLED 15 SP2
Console de surveillance et maître de licence	Effectue une surveillance centralisée de l'ensemble du déploiement Splunk et gère les licences Splunk	1	16 cœurs	32 Go de RAM	SLED 15 SP2

Validation des performances du magasin à distance SmartStore

Dans cette validation des performances, nous avons configuré le cache SmartStore dans le stockage local sur tous les indexeurs pour 10 jours de données. Nous avons permis à `maxDataSize=auto` (taille du bucket de 750 Mo) dans le gestionnaire de cluster Splunk et a poussé les modifications vers tous les indexeurs. Pour mesurer les performances de téléchargement, nous avons ingéré 10 To par jour pendant 10 jours et avons transféré tous les buckets chauds vers des buckets chauds en même temps et capturé le débit maximal et moyen par instance et à l'échelle du déploiement à partir du tableau de bord de la console de surveillance SmartStore.

Cette image montre les données ingérées en une journée.



Nous avons exécuté la commande suivante à partir du cluster master (le nom de l'index est `eventgen-test`). Nous avons ensuite capturé le débit de téléchargement maximal et moyen par instance et à l'échelle du déploiement via les tableaux de bord de la console de surveillance SmartStore.

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

Le maître du cluster dispose d'une authentification sans mot de passe pour tous les indexeurs (rtp-idx0001…rtp-idx0018).

Pour mesurer les performances de téléchargement, nous avons expulsé toutes les données du cache en exécutant la CLI d'expulsion deux fois à l'aide de la commande suivante.

Nous avons exécuté la commande suivante à partir du cluster master et exécuté la recherche à partir de la tête de recherche sur 10 jours de données du magasin distant de StorageGRID. Nous avons ensuite capturé le débit de téléchargement maximal et moyen par instance et à l'échelle du déploiement via les tableaux de bord de la console de surveillance SmartStore.

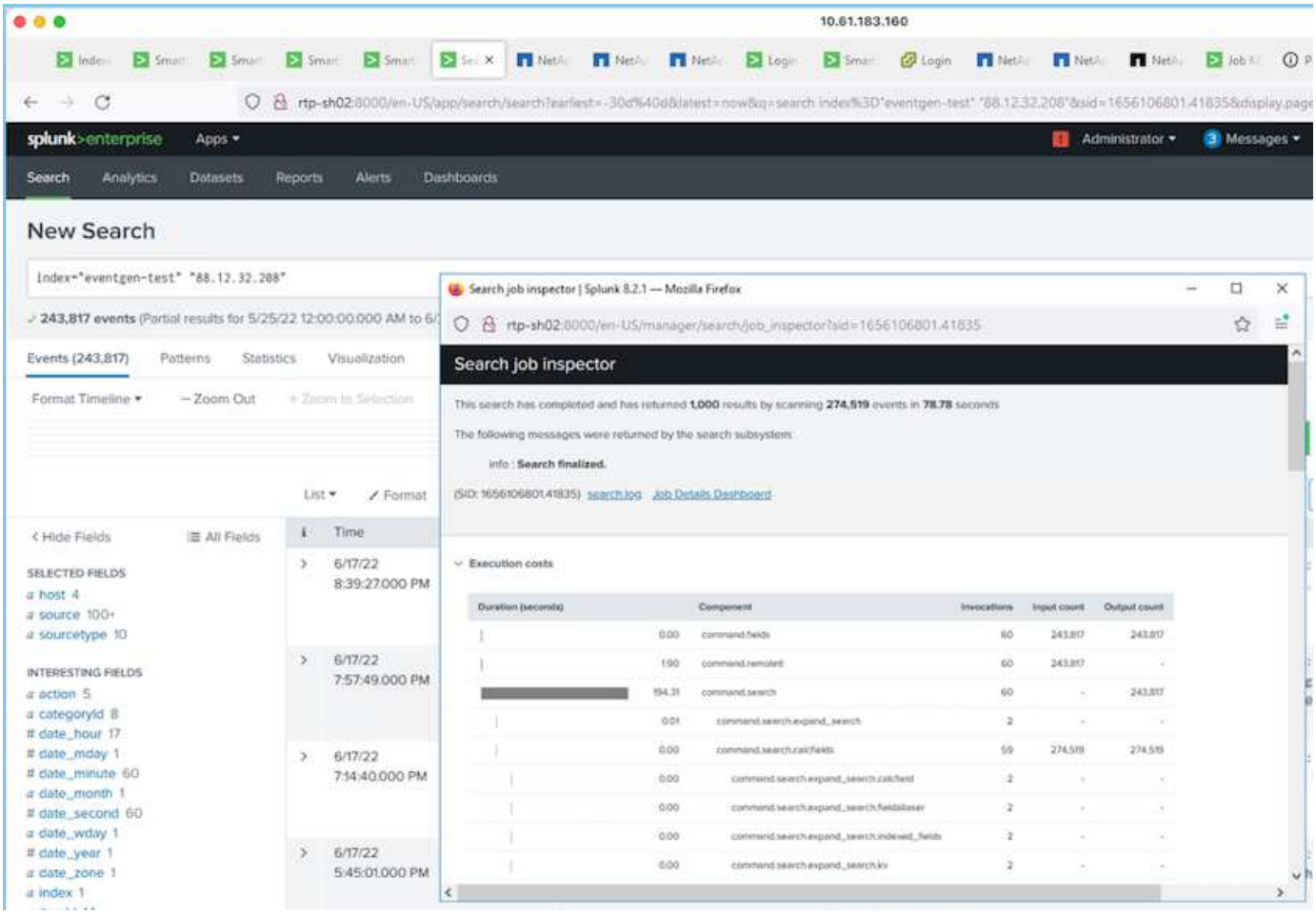
```
for i in rtp-idx0001 rtp-idx0002 rtp-idx0003 rtp-idx0004 rtp-idx0005 rtp-
idx0006 rtp-idx0007 rtp-idx0008 rtp-idx0009 rtp-idx0010 rtp-idx0011 rtp-
idx0012 rtp-idx0013 rtp-idx0014 rtp-idx0015 rtp-idx0016 rtp-idx0017 rtp-
idx0018 ; do  ssh $i " hostname;  date; /opt/splunk/bin/splunk _internal
call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

Les configurations de l'indexeur ont été poussées depuis le maître du cluster SmartStore. Le maître du cluster avait la configuration suivante pour l'indexeur.

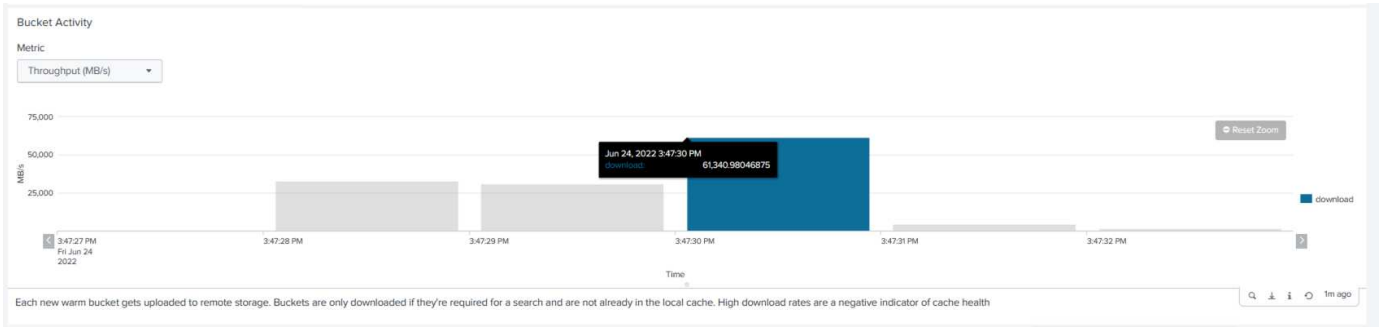
```
Rtp-cm01:~ # cat /opt/splunk/etc/master-apps/_cluster/local/indexes.conf
[default]
maxDataSize = auto
#defaultDatabase = eventgen-basic
defaultDatabase = eventgen-test
hotlist_recency_secs = 864000
repFactor = auto
[volume:remote_store]
storageType = remote
path = s3://smartstore2
remote.s3.access_key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

Nous avons exécuté la requête de recherche suivante sur la tête de recherche pour collecter la matrice de performances.



Nous avons collecté les informations de performance du cluster master. La performance maximale était de 61,34 Gbit/s.

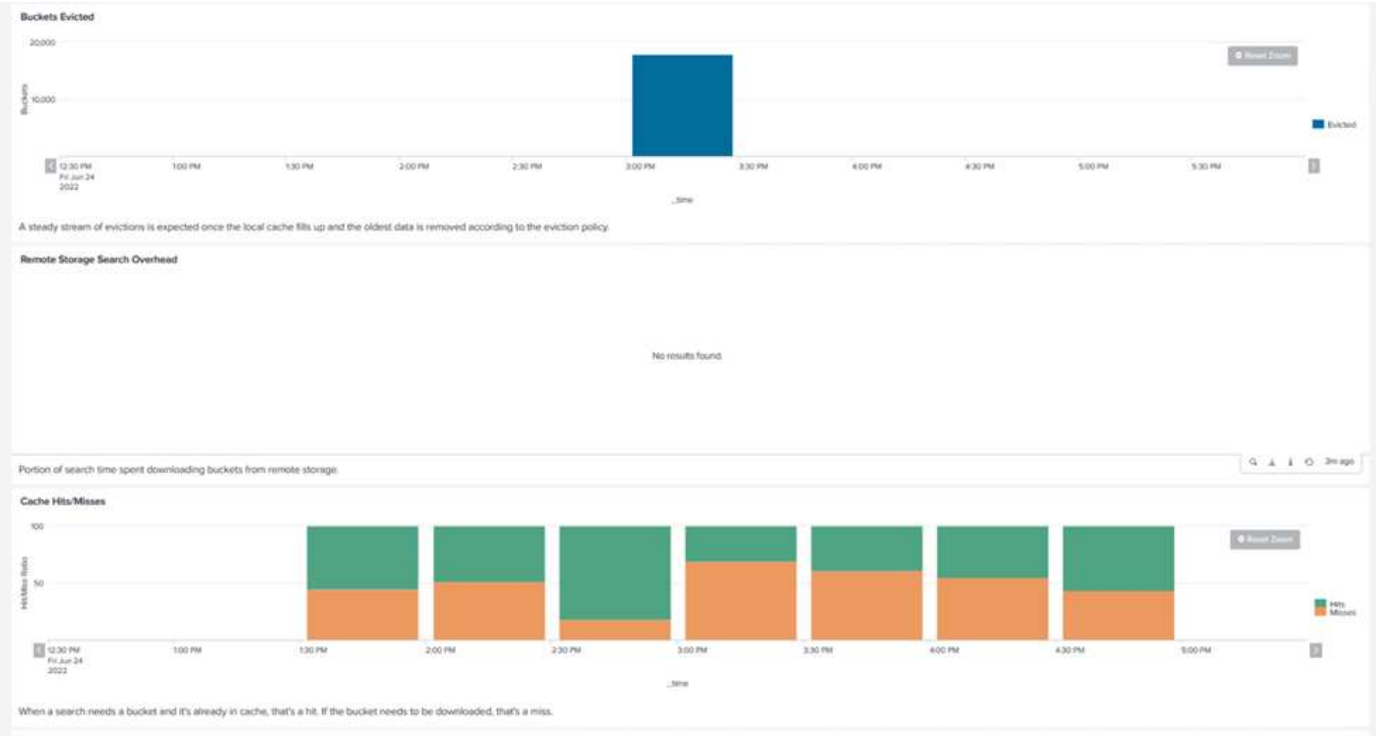


La performance moyenne était d’environ 29 Gbit/s.

Performances de StorageGRID

Les performances de SmartStore sont basées sur la recherche de modèles et de chaînes spécifiques à partir de grandes quantités de données. Dans cette validation, les événements sont générés à l'aide de "Eventgen" sur un index Splunk spécifique (eventgen-test) via la tête de recherche, et la requête va à StorageGRID pour la plupart des requêtes. L'image suivante montre les succès et les échecs des données de requête. Les données de hits proviennent du disque local et les données d'échecs proviennent du contrôleur StorageGRID .

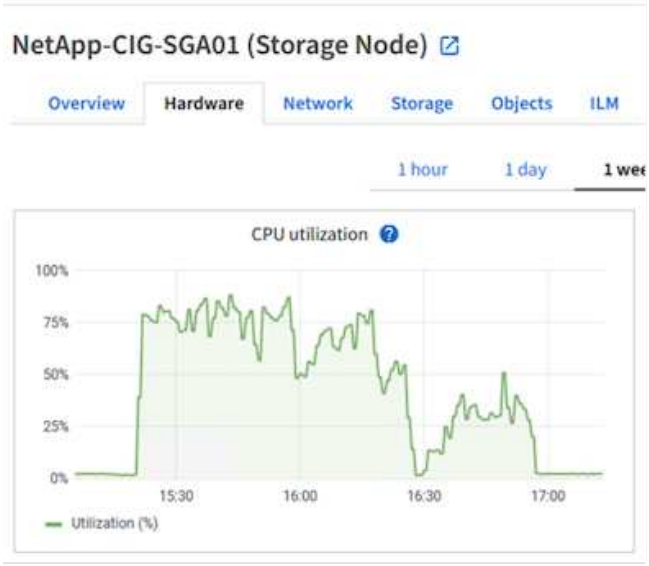
La couleur verte montre les données de hits et la couleur orange montre les données d'échecs.

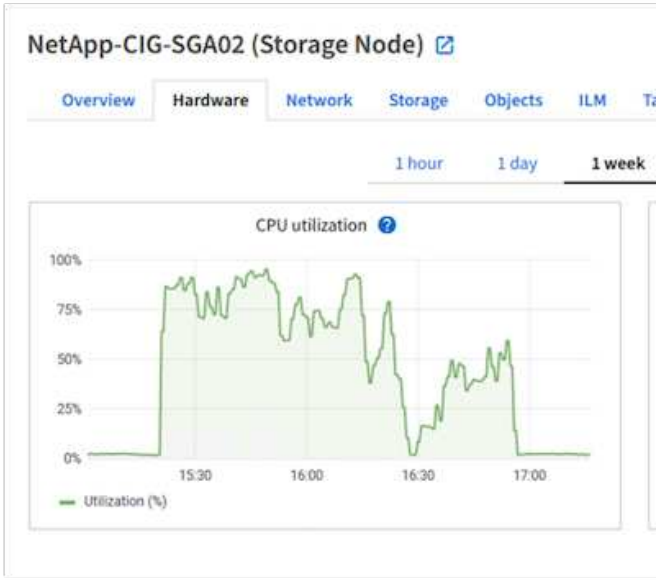


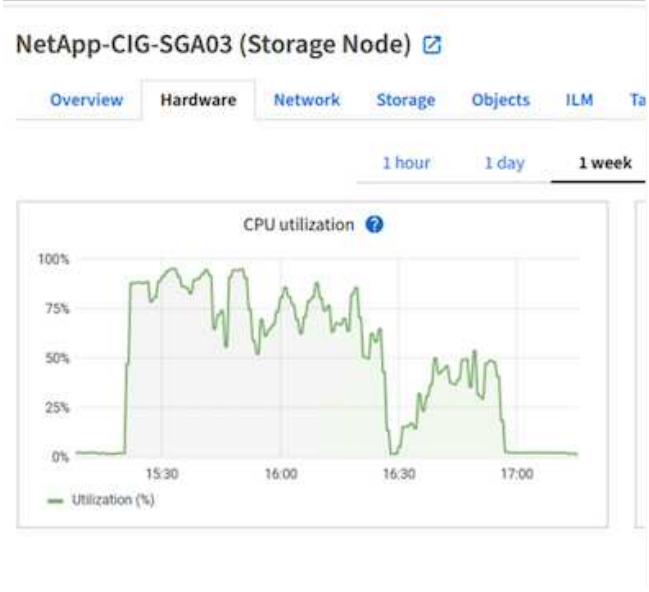
Lorsque la requête s'exécute pour la recherche sur StorageGRID, le temps de récupération S3 à partir de StorageGRID est indiqué dans l'image suivante.

Utilisation du matériel StorageGRID

L'instance StorageGRID dispose d'un équilibreur de charge et de trois contrôleurs StorageGRID . L'utilisation du processeur pour les trois contrôleurs est comprise entre 75 % et 100 %.







SmartStore avec contrôleur de stockage NetApp - avantages pour le client

- **Découplage du calcul et du stockage.** Splunk SmartStore dissocie le calcul et le stockage, ce qui vous aide à les faire évoluer indépendamment.
- **Données à la demande.** SmartStore rapproche les données du calcul à la demande et offre une élasticité de calcul et de stockage ainsi qu'une rentabilité pour obtenir une conservation des données plus longue à grande échelle.
- **Conforme à l'API AWS S3.** SmartStore utilise l'API AWS S3 pour communiquer avec le stockage de restauration, qui est un magasin d'objets compatible AWS S3 et S3 API tel que StorageGRID.
- **Réduit les besoins et les coûts de stockage.** SmartStore réduit les besoins de stockage des données anciennes (chaudes/froides). Une seule copie des données est nécessaire, car le stockage NetApp assure la protection des données et prend en charge les pannes et la haute disponibilité.
- **Panne matérielle.** Une défaillance de nœud dans un déploiement SmartStore ne rend pas les données inaccessibles et permet une récupération de l'indexeur beaucoup plus rapide en cas de défaillance matérielle ou de déséquilibre des données.
- Cache sensible aux applications et aux données.

- Ajoutez-supprimez des indexeurs et configurez-désinstallez des clusters à la demande.
- Le niveau de stockage n'est plus lié au matériel.

Conclusion

Splunk Enterprise est la solution SIEM leader du marché qui génère des résultats pour les équipes de sécurité, d'informatique et de DevOps. L'utilisation de Splunk a considérablement augmenté dans les organisations de nos clients. Il est donc nécessaire d'ajouter davantage de sources de données tout en conservant les données pendant une période plus longue, sollicitant ainsi l'infrastructure Splunk.

La combinaison de Splunk SmartStore et de NetApp StorageGRID est conçue pour fournir une architecture évolutive permettant aux organisations d'obtenir des performances d'ingestion améliorées avec le stockage d'objets SmartStore et StorageGRID et une évolutivité accrue pour un environnement Splunk dans plusieurs régions géographiques.

Où trouver des informations supplémentaires

Pour en savoir plus sur les informations décrites dans ce document, consultez les documents et/ou sites Web suivants :

- "Ressources de documentation NetApp StorageGRID"
- "Documentation produit NetApp"
- "Documentation de Splunk Enterprise"
- "Splunk Enterprise À propos de SmartStore"
- "Manuel de déploiement distribué Splunk Enterprise"
- "Splunk Enterprise Gestion des indexeurs et des clusters d'indexeurs"

Informations sur le copyright

Informations sur les marques commerciales