

ARCHITECTURES VALIDÉES PAR SPLUNK

Sommaire

Introduction.....	2
Structure du document.....	2
Pourquoi utiliser les architectures validées par Splunk.....	2
Bases fondamentales des architectures validées par Splunk.....	3
Qu'attendre des architectures validées par Splunk	4
Rôles et responsabilités.....	4
Présentation du processus de sélection des architectures validées par Splunk	5
Étape 1a : définir vos exigences en indexation et en recherche.....	6
Étape 2a : choisir une topologie d'indexation et de recherche.....	12
Étape 1b : définir vos exigences en collecte de données	23
Étape 2b : choisir vos composants de collecte de données	28
Étape 3 : appliquer les principes de conception et les bonnes pratiques.....	41
Récapitulatif et étapes suivantes.....	51
Étapes suivantes	51
Annexe.....	52
Annexe A : détails des bases fondamentales des AVS.....	52
Annexe B : composants de topologie.....	53

Introduction

Les architectures validées par Splunk (AVS) sont des architectures de référence éprouvées pour des déploiements Splunk stables, efficaces et reproductibles. Pour de nombreux clients existants de Splunk, l'adoption et l'expansion ont été rapides, ce qui a entraîné certaines difficultés lors des tentatives d'élargissement de leur déploiement. Parallèlement à cela, les nouveaux clients Splunk sont de plus en plus demandeurs en consignes et architectures certifiées, le but étant d'acquiescer la certitude que leur déploiement initial repose sur des bases solides. Les AVS ont été élaborées pour répondre à ces besoins croissants.

Que vous soyez client depuis peu ou depuis longtemps, les AVS vous aideront à mettre en place un environnement plus facile à entretenir et plus simple à dépanner. Les AVS sont conçues pour vous apporter les meilleurs résultats possibles tout en minimisant votre coût total de possession. De plus, l'ensemble de votre base Splunk reposera sur une architecture reproductible qui vous permettra d'élargir votre déploiement au fil de l'évolution de vos besoins.

Les AVS proposent des options de topologie répondant à un large éventail de spécifications organisationnelles : vous pourrez ainsi aisément comprendre et identifier la topologie conforme à vos exigences. Le processus de sélection des architectures validées par Splunk vous aidera à associer vos exigences à la topologie la mieux adaptée. Si vous débutez avec Splunk, nous vous recommandons de mettre en œuvre une architecture validée pour votre déploiement initial. Si vous êtes client depuis longtemps, nous vous recommandons d'étudier la possibilité d'aligner votre déploiement sur une topologie d'architecture validée. À moins que vous n'ayez des exigences uniques qui rendent indispensable l'élaboration d'une architecture personnalisée, il est plus que probable qu'une architecture validée répondra à vos exigences tout en restant économique.

Ce livre blanc présente les AVS. Vous y trouverez les ressources dont vous avez besoin pour mener à bien le processus de sélection d'AVS : questionnaire sur les besoins, schémas de topologie de déploiement, principes de conception et consignes générales.

S'il vous faut de l'aide pour déployer une architecture validée par Splunk, contactez les [Services professionnels Splunk](https://www.splunk.com/en_us/support-and-services/splunk-services.html) (https://www.splunk.com/en_us/support-and-services/splunk-services.html).

Structure du document

Les AVS sont réparties en trois grands domaines de contenu :

1. Topologies d'indexation et de recherche
2. Composants d'architecture de collecte de données
3. Principes de conception et bonnes pratiques

L'indexation et la recherche couvrent la couche d'architecture qui fournit les capacités fondamentales d'indexation et de recherche d'un déploiement Splunk. La section consacrée aux composants de collecte de données vous guide dans le choix du mécanisme de collecte le mieux adapté à vos exigences.

Les Principes de conception et Bonnes pratiques s'appliquent à l'ensemble de votre architecture et vous permettront de faire les bons choix lorsque vous établirez les détails de votre déploiement.

Pourquoi utiliser les architectures validées par Splunk

La mise en œuvre d'une architecture validée vous permettra de concevoir votre déploiement Splunk avec plus de certitude. Les AVS relèvent des défis courants rencontrés par les entreprises, notamment dans les domaines suivants :

Performances

- Les entreprises veulent constater des améliorations des performances et de la stabilité.

Complexité

- Les sociétés tombent souvent dans les écueils des déploiements personnalisés, surtout lorsque leur expansion a été trop rapide ou organique. Dans ces situations, il est courant qu'un degré de complexité inutile intervienne dans l'environnement. Cette complexité peut devenir un obstacle majeur lors d'une tentative d'élargissement.

Efficacité

- Pour maximiser les avantages tirés d'un déploiement Splunk, les entreprises doivent améliorer l'efficacité de leurs opérations et réduire le délai de génération de valeur.

Coût

- Les entreprises cherchent des moyens de réduire leur coût total de possession (TCO) tout en satisfaisant toutes leurs exigences.

Agilité

- Les sociétés doivent s'adapter aux changements au fil de leur croissance.

Maintenance

- Il est souvent nécessaire d'optimiser l'environnement pour réduire les efforts de maintenance.

Évolutivité

- Les entreprises doivent avoir la possibilité d'évoluer de façon efficace et transparente.

Vérification

- Les parties prenantes de l'entreprise veulent être certaines que leur déploiement Splunk reposera sur les bonnes pratiques.

Bases fondamentales des architectures validées par Splunk

Les architectures validées par Splunk reposent sur les bases fondamentales suivantes. Pour plus d'informations sur les bases fondamentales de conception, consultez l'Annexe A figurant dans la suite de ce document.

DISPONIBILITÉ	PERFORMANCES	ÉVOLUTIVITÉ	SÉCURITÉ	GÉRABILITÉ
Le système est continuellement opérationnel et capable de se rétablir en cas de panne ou de perturbation prévue ou imprévue.	Le système peut maintenir un niveau de service optimal dans divers schémas d'utilisation.	Le système est conçu pour pouvoir évoluer sur toutes les couches, ce qui vous permet de gérer efficacement le volume croissant d'applicatifs .	Le système est pensé pour protéger les données, les configurations et les actifs tout en continuant à générer de la valeur.	Le système est contrôlable et gérable de manière centralisée sur toutes les couches .

Ces bases fondamentales soutiennent directement le service **Gestion et assistance de la plateforme** du modèle Centra d'excellence Splunk.

Qu'attendre des architectures validées par Splunk

Notez que les AVS n'incluent pas les technologies de déploiement et les méthodes de dimensionnement. La raison en est la suivante :

- Les technologies de déploiement telles que les systèmes d'exploitation et le matériel des serveurs sont considérées comme des choix de déploiement dans le contexte des AVS. Ces choix peuvent varier selon les clients et il n'est pas possible de se contenter de généraliser.
- Le dimensionnement du déploiement nécessite d'évaluer le volume d'ingestion de données, les types de données, les volumes de recherche et les scénarios d'utilisation de recherche, qui sont majoritairement propres à chaque client et n'ont pas de réel impact sur les bases fondamentales de l'architecture de déploiement à proprement parler. Il existe des outils de dimensionnement qui vous faciliteront ce processus une fois que vous aurez établi votre architecture de déploiement. [Splunk Storage Sizing](https://splunk-sizing.appspot.com/) (<https://splunk-sizing.appspot.com/>) est l'un de ces outils.

Les AVS <u>fournissent</u> les éléments suivants :	Les AVS <u>ne fournissent pas</u> les éléments suivants :
<ul style="list-style-type: none"> ✓ Options de déploiement en cluster et sans cluster. ✓ Schémas de l'architecture de référence. ✓ Consignes de sélection de l'architecture la mieux adaptée. ✓ Recommandations propres à chaque couche. ✓ Bonnes pratiques pour le développement de votre déploiement Splunk 	<ul style="list-style-type: none"> ✗ Choix de déploiement (système d'exploitation, physique/virtuel/cloud, etc.). ✗ Dimensionnement du déploiement. ✗ Validation prescriptive de votre architecture. Remarque : les AVS fournissent des recommandations et consignes vous permettant de prendre, à terme, la bonne décision pour votre entreprise. ✗ Suggestion de topologie pour chaque scénario de déploiement possible. Dans certains cas, des facteurs spécifiques peuvent imposer l'élaboration d'une architecture personnalisée. Les experts de Splunk sont à votre disposition pour vous aider dans le développement des solutions dont vous avez besoin. Si vous êtes déjà client, contactez votre équipe chargée de clientèle Splunk. Si vous débutez avec Splunk, vous pouvez nous joindre ici (https://www.splunk.com/en_us/talk-to-sales.html).

Rôles et responsabilités

Les architectures validées par Splunk répondent très directement aux préoccupations des décideurs et administrateurs. Les architectes, consultants, administrateurs Splunk et prestataires de services gérés doivent tous être impliqués dans le processus de sélection de l'AVS. Vous trouverez ci-dessous une description de chacun de ces rôles :

Rôle	Description
Architectes de l'entreprise	Chargés d'élaborer une architecture de déploiement Splunk répondant aux besoins de l'entreprise.
Consultants	Chargés de fournir les services d'architecture, de conception et de déploiement de Splunk.

Rôle	Description
Ingénieurs Splunk	Chargés de gérer le cycle de vie de Splunk.
Prestataires de services gérés	Entités qui déploient et exécutent Splunk en tant que service pour leurs clients.

Présentation du processus de sélection des architectures validées par Splunk

Le processus de sélection des architectures validées par Splunk vous aidera à identifier l'architecture la plus simple et la plus standard répondant à l'ensemble des besoins de votre entreprise.



Étapes du processus de sélection	Objectifs	Considérations
Étape 1 : définir les exigences en : a) Indexation et recherche b) Mécanismes de collecte de données	<i>Définissez les exigences.</i>	<ul style="list-style-type: none"> Les décideurs, les parties prenantes et les administrateurs doivent travailler ensemble identifier et définir les exigences de votre entreprise. Si un déploiement est déjà en place, vous pouvez évaluer votre architecture actuelle pour savoir ce qu'apporterait l'adoption d'un modèle validé. <p><i>Vous trouverez un questionnaire permettant de définir vos exigences à l'étape 1 plus bas.</i></p>
Étape 2 : choisir une topologie pour : A) Indexation et recherche b) Chaque mécanisme de collecte de données	<i>Choisissez une topologie répondant aux exigences identifiées.</i>	<ul style="list-style-type: none"> Vous allez choisir la topologie répondant le mieux à vos exigences. Privilégiez la simplicité et restez conforme à l'AVS pour profiter de ses avantages en termes d'évolutivité simplifiée. <p><i>Pour obtenir des schémas et descriptions des options de topologie, consultez l'étape 2 ci-dessous.</i></p>

Étape 3 : appliquer les principes de conception et les bonnes pratiques	<i>Hiérarchisez vos principes de conception et passez en revue les bonnes pratiques de déploiement propres à chaque couche.</i>	<ul style="list-style-type: none"> • Chaque principe de conception renforce une ou plusieurs bases fondamentales des architectures validées par Splunk. • Vous allez hiérarchiser les principes de conception conformément aux exigences de votre entreprise. • Les recommandations propres à chaque couche vous guideront dans le déploiement de votre topologie. <p><i>Pour plus de détails sur les principes de conception, consultez l'étape 3 ci-dessous.</i></p>
--	---	---

Étape 1a : définir vos exigences en indexation et en recherche

Pour choisir la topologie de déploiement appropriée, vous devrez procéder à un examen approfondi de vos exigences. Une fois que vous aurez établi votre cahier des charges, vous pourrez choisir l'approche la plus simple et la plus rentable pour le déploiement de Splunk. Vous trouverez ci-dessous un questionnaire qui vous aidera à définir des domaines d'exigences clés concernant les couches d'indexation et de recherche de votre déploiement.

Le questionnaire de spécification se concentre sur les aspects qui ont un impact direct sur la topologie de votre déploiement. Par conséquent, nous vous recommandons vivement de consigner les réponses aux questions avant de choisir une topologie à l'étape suivante.

Aspects à prendre en compte

Examinez vos scénarios d'utilisation

Lorsque vous définissez vos exigences, vous devez réfléchir aux scénarios d'utilisation prévus de votre infrastructure Splunk. Par exemple, la topologie du scénario d'utilisation d'un département DevOps est souvent plus simple que celle d'un scénario d'utilisation stratégique (mais ce n'est pas systématique). Vous devez étudier avec attention les scénarios d'utilisation impliquant :

- Recherche
- Disponibilité
- Des exigences de conformité (particulièrement crucial si vous devez à tout moment assurer 100 % de fidélité et de disponibilité des données)
- D'autres scénarios d'utilisation propres à votre entreprise

Selon vos scénarios d'utilisation, votre déploiement devra peut-être inclure des caractéristiques architecturales supplémentaires.

Pensez à la croissance future

Vous devrez réfléchir à vos exigences immédiates pour établir votre cahier des charges. Mais vous devez également réfléchir à la croissance et à l'évolutivité de votre déploiement. Faire évoluer votre déploiement peut impliquer des dépenses, du personnel supplémentaire ou d'autres ressources qu'il peut être intéressant de planifier dès maintenant.

Catégories de topologies

Vous trouverez ci-dessous les détails des catégories de topologies. Ces catégories sont utilisées dans le questionnaire ci-dessous. Vous trouverez également des références à ces catégories dans les étapes suivantes du processus de sélection de l'AVS.

Catégories de couches d'indexation

Code de catégorie	Explication
S	La catégorie « S » désigne l'indexeur d'un déploiement Splunk à un seul serveur.
D	La catégorie « D » désigne une couche d'indexeurs distribués comprenant au moins 2 indexeurs.
C	La catégorie « C » désigne une couche d'indexeurs en cluster (avec réplication obligatoire des données).
M	La catégorie « M » désigne une couche d'indexeurs distribués avec de multiples sites.

Catégories de couches de recherche

Code de catégorie	Explication
1	La catégorie « 1 » indique qu'une seule search head peut respecter les exigences.
2	La catégorie « 2 » indique qu'il faut plusieurs search heads pour respecter les exigences.
3	La catégorie « 3 » indique qu'il faut un cluster de search heads pour respecter les exigences.
4	La catégorie « 4 » indique qu'il faut un cluster de search heads couvrant de multiples sites (un SHC « étendu ») pour respecter les exigences.
+10	La catégorie « +10 » indique qu'une search head (ou un cluster) dédiée est nécessaire pour prendre en charge l'application Enterprise Security. Ajoutez 10 à la catégorie de topologie de la couche de recherche et lisez attentivement la description de la topologie pour connaître les spécifications propres à cette application.

Questionnaire 1 : définition des exigences en couches d'indexation et de recherche

♦ Consultez la légende ci-dessus pour comprendre les codes des catégories de topologies. Si vous répondez « oui » à plusieurs questions, utilisez le code de catégorie de la question présentant le numéro le plus grand.

#	Question	Considérations	Impact sur la topologie	Catégorie de topologie de la couche d'indexation	Catégorie de topologie de la couche de recherche
1	Prévoyez-vous un volume d'ingestion quotidien inférieur à 300 Go/jour ?	Pensez à la croissance à court terme (sur 6 à 12 mois) de l'ingestion quotidienne	Candidat à un déploiement à un seul serveur, sous réserve des réponses aux questions concernant la disponibilité	S	1
2	Vous faut-il une haute disponibilité de l'indexation et de la collecte de données ?	Si vous ne prévoyez pas d'utiliser Splunk pour des scénarios de surveillance nécessitant une ingestion continue des données, une interruption momentanée du flux de données entrant peut être acceptable, à condition qu'aucune donnée de log ne soit perdue.	Nécessite un déploiement distribué pour prendre en charge l'ingestion en continu	D	1
3	En partant du principe qu'une search head est disponible pour exécuter une recherche : Vos données doivent-elles être intégralement interrogeables ? Autrement dit, l'exhaustivité des résultats de recherche ne peut-elle en aucun cas être affectée ?	Si votre scénario d'utilisation implique un calcul de métriques de performances et une surveillance générale à l'aide de fonctions d'agrégation, par exemple, une seule interruption sur un indexeur n'aura sans doute pas un impact décisif sur le calcul de statistiques portant sur un grand nombre d'événements. Si votre scénario d'utilisation implique une surveillance de sécurité et une détection des menaces, la moindre tache aveugle dans les résultats de recherche peut avoir des conséquences indésirables.	Nécessite des indexeurs en cluster avec un facteur de réplication d'au moins deux (2). Remarque : Bien qu'un facteur de réplication de 2 offre une protection minimale contre la défaillance d'un seul nœud d'indexeur, le facteur de réplication recommandé (et appliqué par défaut) est de 3.	C	1
4	Exploitez-vous plusieurs datacenters et vous avez	Les impératifs de récupération en cas de sinistre peuvent imposer un fonctionnement	Le fonctionnement en continu nécessitera des clusters d'indexeurs	M	2

#	Question	Considérations	Impact sur la topologie	Catégorie de topologie de la couche d'indexation	Catégorie de topologie de la couche de recherche
	besoin d'une restauration automatique de votre environnement Splunk en cas de panne d'un datacenter ?	continu sur deux installations (active/active) ou prescrire des objectifs RTO/RPO pour une récupération manuelle.	multisites et au moins deux search heads actives pour garantir le basculement en cas de panne au niveau de la couche d'ingestion/indexation comme au niveau de la couche de recherche.		
5	En prenant l'hypothèse d'une ingestion continue et sans perte de données, la couche de recherche côté utilisateur exige-t-elle une haute disponibilité ?	Si Splunk est utilisé à des fins de surveillance continue en quasi-temps réel, les interruptions de la couche de recherche seront probablement inacceptables. Cela peut être le cas dans d'autres scénarios d'utilisation.	Nécessite des search heads redondantes, et potentiellement un cluster de search heads.	D/C/M	3
6	Devez-vous prendre en charge un grand nombre d'utilisateurs simultanés et/ou une charge importante de recherches planifiées ?	La présence de plus de 50 utilisateurs/recherches simultanés nécessite généralement une expansion horizontale de la couche de recherche.	Peut nécessiter une topologie reposant sur un cluster de search heads dans la couche de recherche.	D/C/M	3
7	Dans un environnement comprenant plusieurs datacenters, avez-vous besoin que les artefacts utilisateur (recherches, tableaux de bord et autres objets de connaissance) soient	C'est ce qui déterminera si les utilisateurs bénéficieront d'une expérience à jour et cohérente en cas de défaillance d'un site.	Nécessite un cluster de search heads « étendu » sur les différents sites et une configuration appropriée. Important : Bien qu'un SHC étendu puisse améliorer la disponibilité pour les utilisateurs en cas de défaillance totale d'un site, il est impossible de garantir que tous les	M	4

#	Question	Considérations	Impact sur la topologie	Catégorie de topologie de la couche d'indexation	Catégorie de topologie de la couche de recherche
	synchronisés entre les différents sites ?		artefacts soient répliqués en permanence sur tous les sites. Cela peut avoir des conséquences sur certaines applications qui reposent sur des artefacts cohérents et à jour, comme l'application Splunk dédiée à la sécurité des entreprises. Le clustering de search heads ne suffit pas à lui seul à fournir une solution de récupération complète. Les autres avantages des SHC restent valables.		
8	Prévoyez-vous de déployer l'application Splunk dédiée à la sécurité des entreprises (ES) ?	Veillez bien <u>lire et comprendre</u> les limitations propres à l'application Splunk dédiée à la sécurité des entreprises, qui sont documentées pour chaque topologie.	ES nécessite un environnement de search head dédiée (autonome ou en cluster).	D/C/M	+10
9	Votre environnement est-il distribué sur différentes régions, et donc soumis à des réglementations sur la protection des données ?	Certains pays n'autorisent pas que les données générées sur leur territoire quittent les systèmes locaux.	Ces réglementations empêchent le déploiement d'une couche d'indexation centralisée et imposent le développement d'une architecture personnalisée, au moyen d'une collaboration entre Splunk ou le Partenaire et le client, et en examinant de façon approfondie les détails du déploiement en question. En d'autres termes, aucune AVS	Personnalisé	Personnalisé

#	Question	Considérations	Impact sur la topologie	Catégorie de topologie de la couche d'indexation	Catégorie de topologie de la couche de recherche
			ne répond à ce cahier des charges.		
10	Avez-vous mis en place des politiques de sécurité très restrictives qui empêchent la cohabitation de sources de données de log spécifiques sur des serveurs/ indexeurs partagés ?	D'après certaines politiques d'entreprise, certaines données de log très sensibles ne peuvent être autorisées à cohabiter avec des jeux de données présentant un risque inférieur sur le même système physique ou au sein de la même zone réseau.	Plusieurs environnements d'indexation indépendants sont nécessaires, éventuellement avec une couche de recherche hybride partagée. Cette approche dépasse le champ d'application des AVS et nécessite l'élaboration d'une architecture personnalisée.	Personnalisé	Personnalisé

Comment déterminer votre code de catégorie de topologie

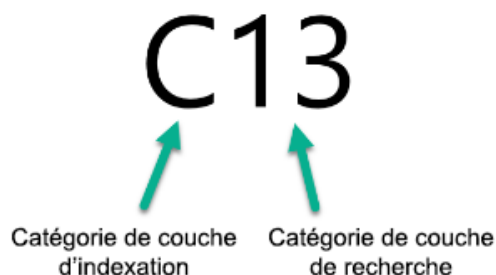
Selon vos réponses au questionnaire de spécification, vous obtiendrez un indicateur combiné de catégories de topologie qui vous permettra d'identifier la meilleure topologie pour vos besoins. Vous trouverez des instructions et des exemples ci-dessous.

Instructions

1. Notez les questions auxquelles vous avez répondu « oui ».
2. Si vous avez répondu « oui » à plusieurs questions, suivez la recommandation de catégorie de la question présentant le numéro le plus grand. Si vous voyez plusieurs options de topologie (D/C/M par exemple), examinez les questions précédentes pour déterminer l'option la plus adaptée.
3. Votre code de catégorie doit commencer par la lettre représentant la couche d'indexation (« C » ou « M » par exemple). Cette lettre est suivie par le numéro représentant la couche de recherche (« 1 » ou « 13 » par exemple).

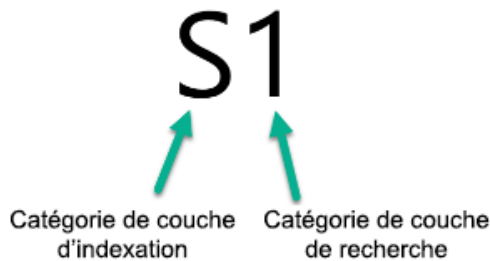
Exemple 1 :

Imaginons que vous ayez répondu « oui » aux questions 3, 5 et 8. Vous obtenez la catégorie de topologie « C13 », ce qui se traduit par une couche d'indexation en cluster comprenant deux clusters de search heads.



Exemple 2 :

Imaginons maintenant que vous ayez seulement répondu « oui » à la question 1. Vous obtenez une catégorie de topologie « S1 », qui recommande un déploiement Splunk à un seul serveur comme topologie idéale.

**Étape 2a : choisir une topologie d'indexation et de recherche**

Les topologies sont généralement réparties en deux groupes : avec et sans cluster. Les déploiements sans cluster requièrent un nombre minimal de composants distincts et présentent d'excellentes caractéristiques d'évolutivité. N'oubliez pas que bien que les déploiements sans cluster aient des avantages moindres en termes de disponibilité et de récupération en cas de sinistre, cette option de déploiement peut tout de même être très adaptée à votre entreprise.

N'oubliez pas que : l'objectif principal du processus de sélection d'AVS consiste à vous permettre de mettre en place ce dont vous avez besoin, sans introduire de composants inutiles.

Remarque

Vous pouvez choisir de déployer une topologie qui apporte des avantages supplémentaires allant au-delà de vos besoins immédiats, mais gardez à l'esprit que cela entraînera vraisemblablement des coûts supplémentaires. De plus, l'introduction d'une complexité supplémentaire est souvent contre-productive sur le plan de l'efficacité opérationnelle.

Remarque importante au sujet des schémas de topologie

Les icônes des schémas de topologie représentent des **rôles fonctionnels Splunk** et n'impliquent pas d'infrastructure dédiée. Veuillez consulter l'Annexe pour savoir quels rôles peuvent cohabiter dans une même infrastructure ou sur un même serveur.

Comment utiliser votre code de catégorie de topologie

Avant de choisir une option de topologie, nous vous recommandons vivement de répondre au questionnaire de spécification pour déterminer votre code de catégorie de topologie. Si vous ne l'avez pas encore fait, veuillez suivre l'étape précédente ci-dessus. Une fois en possession de votre code de catégorie de topologie, vous pourrez identifier l'option de déploiement qui convient le mieux à vos besoins.

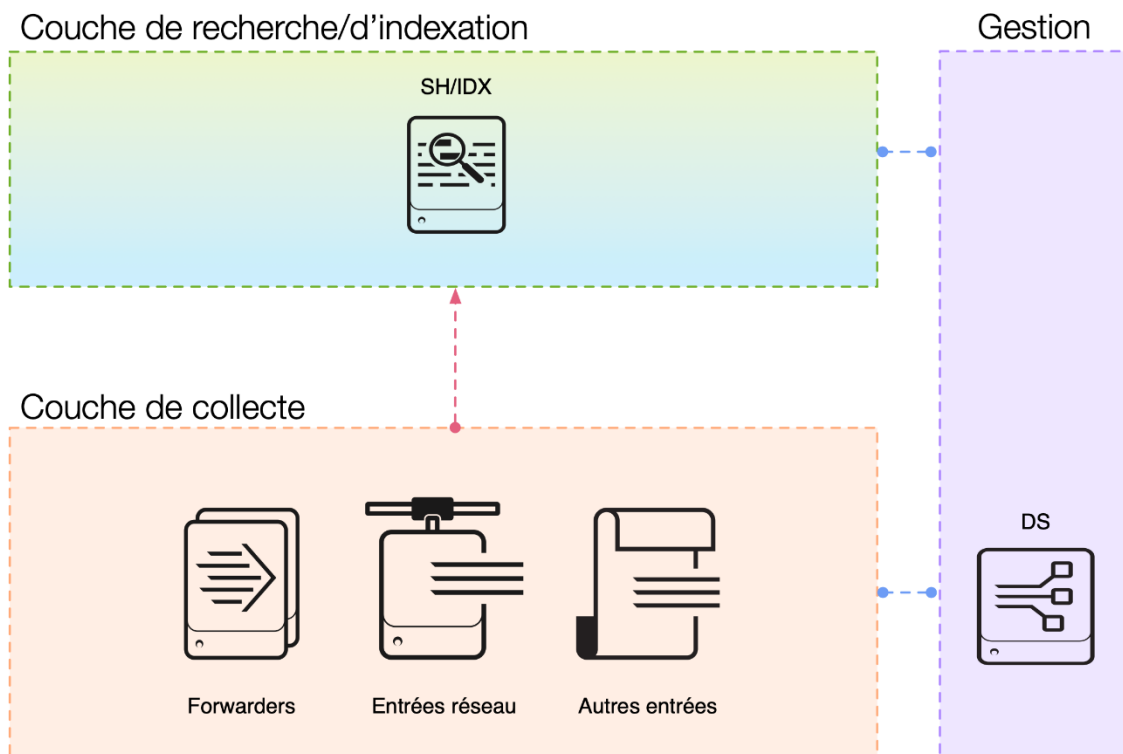
Options de déploiement sans cluster

Vous trouverez ci-dessous les options de topologie suivantes :

Type de déploiement	Code(s) de catégorie de topologie
Déploiement sur un seul serveur	S1
Déploiement sans cluster distribué	D1 / D11

Pour obtenir une explication des composants de topologie, consultez l'Annexe B figurant dans la suite de ce document.

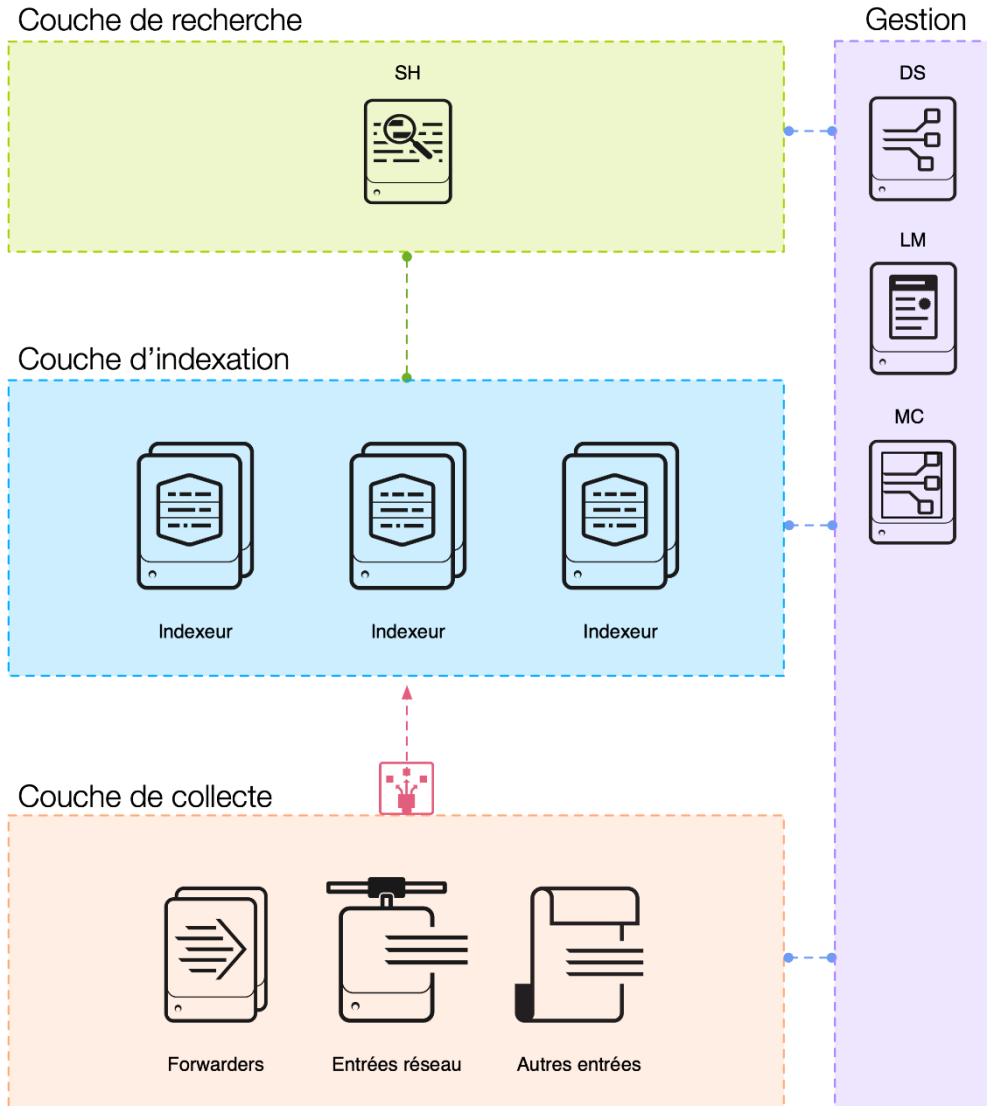
Déploiement sur un seul serveur (S1)



Description du déploiement sur un seul serveur (S1)	Limitations
<p>Cette topologie de déploiement offre une solution très économique si votre environnement répond à l'ensemble des critères suivants : a) il ne vous est pas nécessaire de garantir une haute disponibilité ou la récupération en cas de sinistre pour votre déploiement Splunk ; b) votre volume quotidien d'ingestion de données est inférieur à 300 Go/jour ; c) votre nombre d'utilisateurs est faible et vos scénarios d'utilisation de recherche sont non critiques.</p> <p>Cette topologie est généralement utilisée pour les scénarios d'utilisation non critiques de faible envergure (souvent à l'échelle d'un département). Les environnements de test d'importation de données, les scénarios d'utilisation DevOps de faible envergure, les environnements de test d'applications et d'intégration et autres scénarios similaires sont adaptés à cette topologie.</p>	<ul style="list-style-type: none"> Pas de haute disponibilité pour la recherche et l'indexation Évolutivité limitée par la capacité matérielle (chemin de migration direct vers un déploiement distribué)

Description du déploiement sur un seul serveur (S1)	Limitations
Les principaux avantages de cette topologie sont sa simplicité de gestion, ses bonnes performances de recherche pour les faibles volumes de données et un TCO fixe.	

Déploiement sans cluster distribué (D1 / D11)



Description du déploiement sans cluster distribué (D1 / D11)	Limitations
Vous devez adopter une topologie distribuée dans les situations suivantes : a) si le volume de données à envoyer quotidiennement à Splunk dépasse la capacité d'un déploiement sur un seul serveur ; b) si vous souhaitez/devez assurer une haute disponibilité de l'assimilation des données. Le déploiement de plusieurs indexeurs indépendants vous permettra de faire évoluer votre capacité d'indexation de façon linéaire et accroîtra implicitement la disponibilité de l'ingestion de données.	<ul style="list-style-type: none"> Pas de haute disponibilité pour la couche de recherche. Haute disponibilité limitée pour la couche d'indexation, et la défaillance d'un nœud peut entraîner des résultats de recherche incomplets en cas de recherches dans l'historique.

Description du déploiement sans cluster distribué (D1 / D11)	Limitations
<p>Le TCO augmentera de façon linéaire et prévisible avec l'ajout de nœuds d'indexeurs. L'introduction recommandée du composant Console de surveillance (MC) vous permet de surveiller la santé et la capacité de votre déploiement distribué. De plus, la MC fournit un système d'alerte centralisé qui vous avertit en cas de situation néfaste dans votre déploiement.</p> <p>Les search heads devront être configurées manuellement avec la liste des pairs de recherche disponibles à chaque ajout de nouvel indexeur. Remarque à l'attention des clients ES : si votre code de catégorie est D1 (autrement dit, si vous prévoyez de déployer l'application Splunk dédiée à la sécurité des entreprises), une seule search head dédiée est nécessaire au déploiement de l'application (non représentée dans le schéma de topologie).</p> <p>La couche de collecte doit être configurée en y incluant la liste des indexeurs cibles (via un serveur de déploiement) à chaque nouvel ajout d'indexeur.</p> <p>Cette topologie de déploiement peut évoluer de façon linéaire jusqu'à 1 000 nœuds d'indexeur et ainsi prendre en charge des volumes extrêmes d'ingestion de données et de recherches.</p> <p>Les performances de recherche seront maintenues pour les jeux de données volumineux grâce à l'exécution de recherches en parallèle sur de nombreux indexeurs (association/réduction).</p> <p>Bien que cela ne fasse pas l'objet d'une topologie distincte, il est possible d'utiliser un cluster de search heads pour augmenter la capacité de recherche sur la couche de recherche (voir la couche de recherche de la topologie C3/C13).</p>	

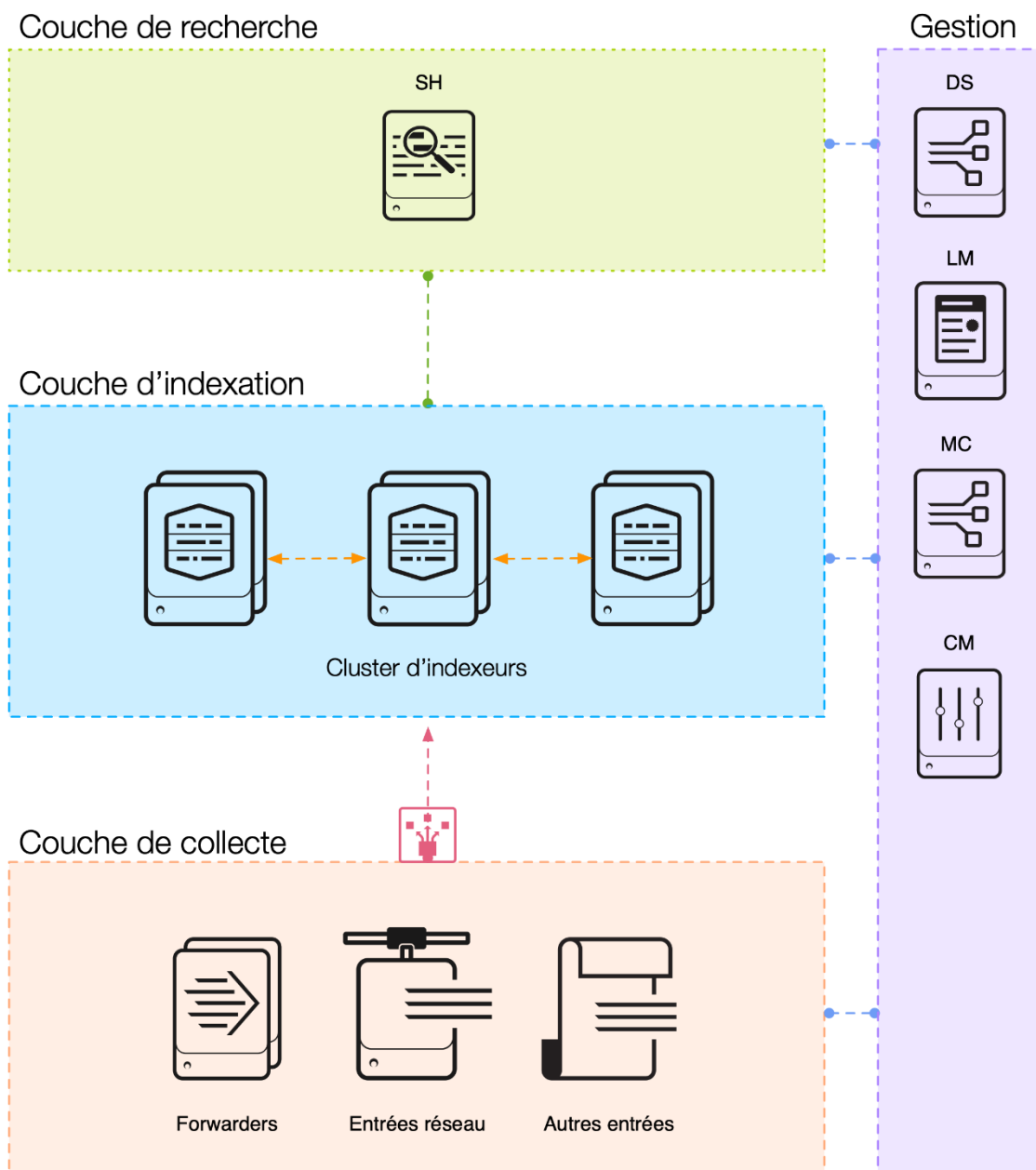
Options de déploiement en cluster

Vous trouverez ci-dessous les options de topologie suivantes :

Type de déploiement	Code(s) de catégorie de topologie
Déploiement en cluster distribué - Un seul site	C1 / C11
Déploiement en cluster distribué + SHC - Un seul site	C3 / C13
Déploiement en cluster distribué - Multisite	M2 / M12
Déploiement en cluster distribué + SHC - Multisite	M3 / M13
Déploiement en cluster distribué + SHC - Multisite	M4 / M14

Pour une explication des composants de topologie, consultez l'Annexe B figurant la suite de ce document.

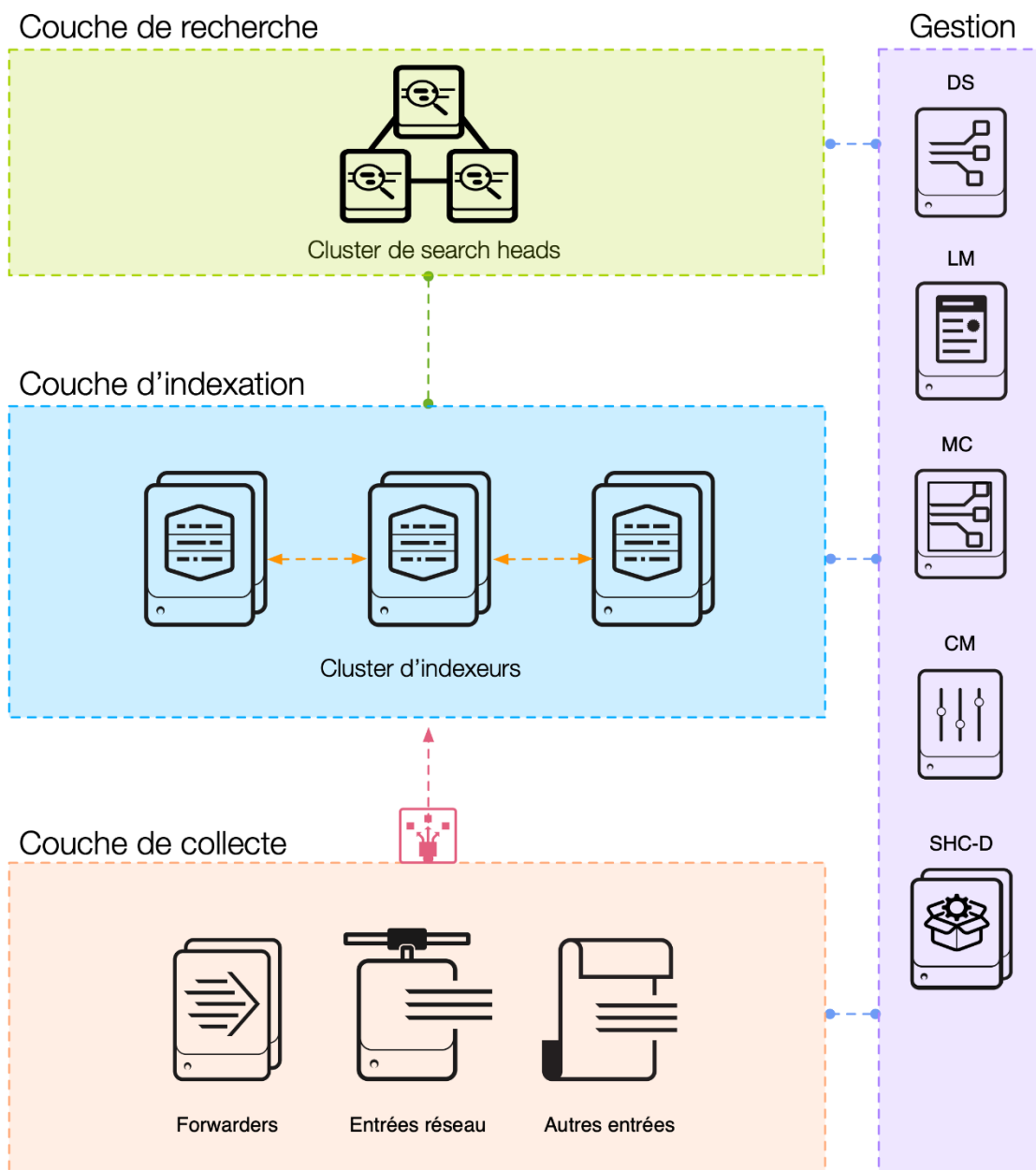
Déploiement en cluster distribué - Un seul site (C1 / C11)



Description du déploiement en cluster distribué - Un seul site (C1 / C11)	Limitations
<p>Cette topologie introduit le clustering d'indexeurs conjoint à une politique de réplication de données configurée de manière appropriée. Cette approche assure la haute disponibilité des données en cas de défaillance d'un nœud pair d'indexeur. Il faut toutefois savoir que cela ne s'applique qu'à la couche d'indexation et ne protège pas contre une défaillance des search heads.</p> <p>Remarque à l'attention des clients ES : si votre code de catégorie est C11 (autrement dit, si vous prévoyez de déployer l'application Splunk dédiée à la sécurité des entreprises), une seule search head dédiée est nécessaire au déploiement de l'application (non représentée dans le schéma de topologie).</p>	<ul style="list-style-type: none"> Pas de haute disponibilité pour la couche de recherche. Le nombre total de buckets uniques du cluster d'indexeurs est limité à 5 millions (V6.6+), pour un total de 15 millions. Aucune capacité de récupération automatique en cas de panne du datacenter.

Description du déploiement en cluster distribué - Un seul site (C1 / C11)	Limitations
<p>Cette topologie nécessite un composant Splunk supplémentaire nommé Maître de clusters (CM). Le CM est responsable de la coordination et de la mise en œuvre de la politique de réplication des données configurée. Le CM joue également le rôle de source d'autorité pour les pairs de cluster (indexeurs) disponibles. La configuration des search heads est plus simple parce que l'on configure le CM plutôt que chaque pair de recherche séparément.</p> <p>Vous avez la possibilité de configurer la couche de transmission pour découvrir d'autres indexeurs via le CM. Cela simplifie la gestion de la couche de transmission.</p> <p>Sachez que les données sont répliquées dans le cluster de façon non déterministe. Vous n'aurez aucun contrôle sur l'emplacement de stockage des copies demandées de chaque événement. De plus, si l'évolutivité est linéaire, la taille totale du cluster est affectée par certaines limites (env. 50 Po de données interrogeables en conditions idéales).</p> <p>Nous vous recommandons de déployer la Console de surveillance (MC) pour surveiller la santé de votre environnement Splunk.</p>	

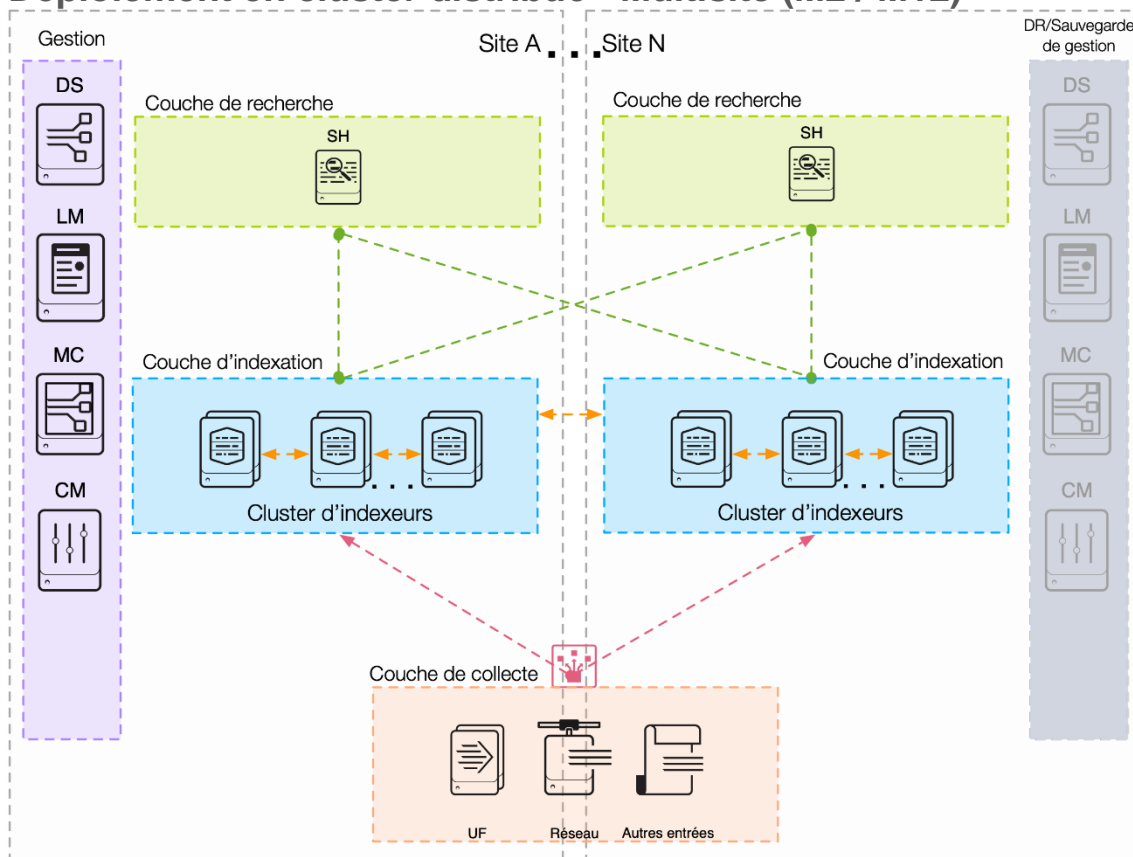
Déploiement en cluster distribué + SHC - Un seul site (C3 / C13)



Description du déploiement en cluster distribué + SHC - Un seul site (C3 / C13)	Limitations
<p>Cette topologie ajoute une évolutivité horizontale et élimine le point de défaillance individuel intervenant au niveau de la couche de recherche. Il faut au moins trois search heads pour déployer un SHC.</p> <p>Pour gérer la configuration du SHC, un autre composant Splunk, appelé Déployeur de cluster de search heads, est nécessaire pour chaque SHC. Ce composant est nécessaire pour déployer les modifications dans les fichiers de configuration du cluster. Le Déployeur de cluster de search heads n'implique aucune spécification de haute disponibilité (aucun rôle runtime).</p> <p>Le SHC fournit le mécanisme qui augmente la capacité de recherche disponible au-delà de ce qu'une seule search head peut fournir. De plus, le SHC permet de répartir la charge de</p>	<ul style="list-style-type: none"> Aucune capacité de récupération en cas de panne du datacenter ES nécessite une SH ou un SHC dédié La gestion d'un déploiement ES dans un SHC est prise en charge mais difficile (implique PS)

Description du déploiement en cluster distribué + SHC - Un seul site (C3 / C13)	Limitations
<p>recherche planifiée dans l'ensemble du cluster. Le SHC permet également un basculement optimal des utilisateurs en cas de défaillance d'une search head.</p> <p>Un équilibreur de charge réseau prenant en charge les sessions persistantes doit également être présent devant chaque membre du SHC, afin de garantir un équilibrage adéquat des utilisateurs dans l'ensemble du cluster.</p> <p>Remarque à l'attention des clients ES : si votre code de catégorie est C13 (autrement dit, si vous prévoyez de déployer l'application Splunk dédiée à la sécurité des entreprises), un cluster de search heads dédié est nécessaire au déploiement de l'application (non représenté dans le schéma de topologie). La couche de recherche peut contenir des search heads en cluster autonomes selon vos besoins en capacité et ceux de votre organisation (cela ne figure pas non plus dans le schéma de la topologie).</p>	<ul style="list-style-type: none"> Le SHC ne peut pas comprendre plus de 100 nœuds.

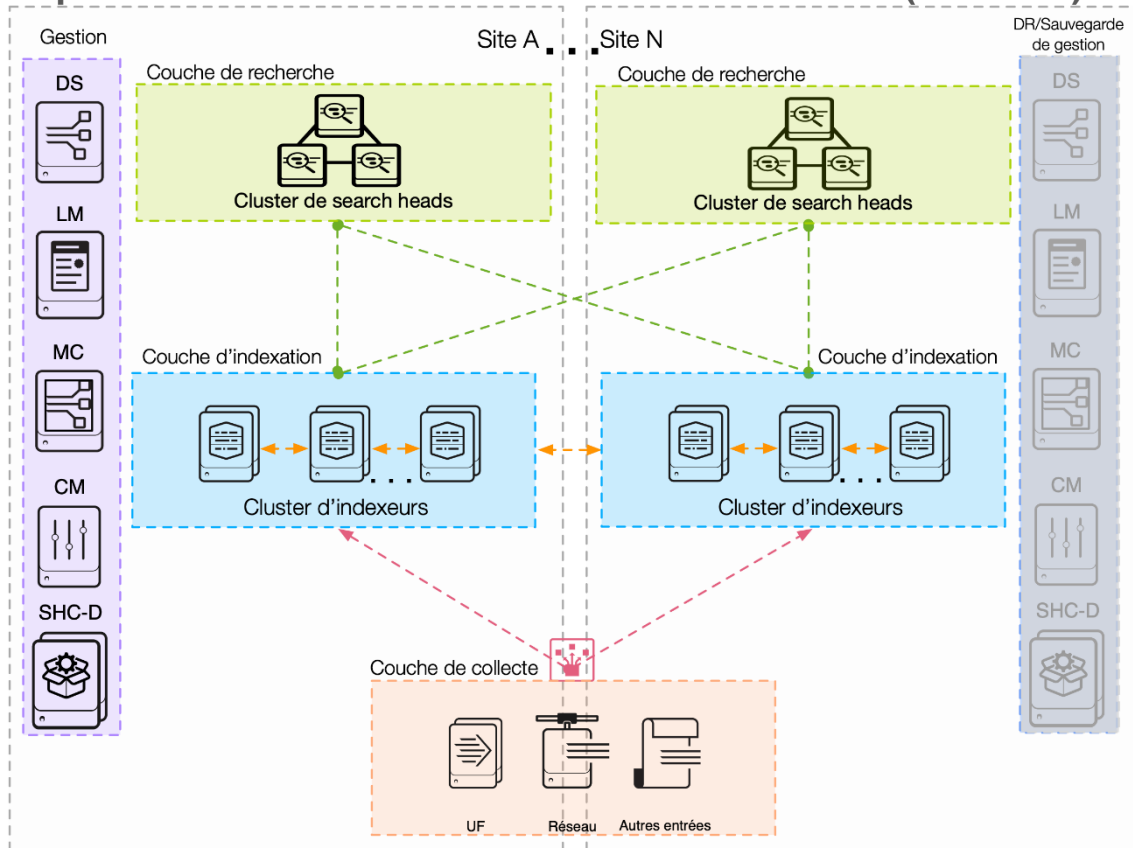
Déploiement en cluster distribué - Multisite (M2 / M12)



Description du déploiement en cluster distribué - Multisite (M2 / M12)	Limitations
<p>Pour assurer la récupération quasi automatique en cas d'événement catastrophique (comme une panne de datacenter), le clustering multisite est l'architecture de déploiement privilégiée. Pour fonctionner correctement, le</p>	<ul style="list-style-type: none"> Aucun partage de la capacité de search head disponible et aucune

Description du déploiement en cluster distribué - Multisite (M2 / M12)	Limitations
<p>cluster multisite nécessite une latence réseau inter-site acceptable, comme spécifié dans la documentation Splunk.</p> <p>Cette topologie vous permet de répliquer les données vers deux groupes de pairs de clusters d'indexeurs ou plus, et ce de façon déterministe. Vous pourrez configurer le facteur de recherche et de réplication de site. Le facteur de réplication de site vous permet de préciser l'emplacement vers lequel les répliques sont envoyées, et garantit ainsi la distribution des données à plusieurs emplacements.</p> <p>Il reste géré par un seul nœud maître de clusters qui doit basculer vers le site de récupération en cas de sinistre.</p> <p>Le clustering multisite assure la redondance des données à des emplacements physiquement séparés, voire répartis dans différentes régions du monde.</p> <p>Concernant l'impératif de disponibilité, les utilisateurs peuvent basculer automatiquement vers le site de récupération. Toutefois, cette topologie ne fournit aucun mécanisme pour synchroniser automatiquement la configuration de la couche de recherche et les artefacts de runtime entre les sites.</p> <p>La capacité disponible en pairs de recherche (indexeurs) des différents sites peut être exploitée pour exécuter des recherches dans un modèle actif/actif. Lorsque la situation le permet, l'affinité des sites peut être configurée pour que les utilisateurs connectés à la search head d'un site spécifique ne puissent interroger que les indexeurs locaux.</p> <p>Remarque à l'attention des clients ES : si votre code de catégorie est M12 (autrement dit, si vous prévoyez de déployer l'application Splunk dédiée à la sécurité des entreprises), une seule search head dédiée est nécessaire au déploiement de l'application (non représentée dans le schéma de topologie). Pour la search head ES, le basculement en cas de panne implique le déploiement d'une search head « fantôme » sur le site de destination ; celle-ci ne sera activée et utilisée qu'en situation de récupération. Veuillez faire appel aux Services professionnels Splunk pour concevoir et déployer un mécanisme de basculement vers un site de secours pour votre déploiement d'Enterprise Security.</p>	<p>réplication des artefacts de recherche entre les sites</p> <ul style="list-style-type: none"> • Une défaillance des fonctions de Gestion doit être traitée en dehors de Splunk en cas de panne d'un site • La latence inter-sites affectant la réplication des index doit être conforme aux limites recommandées

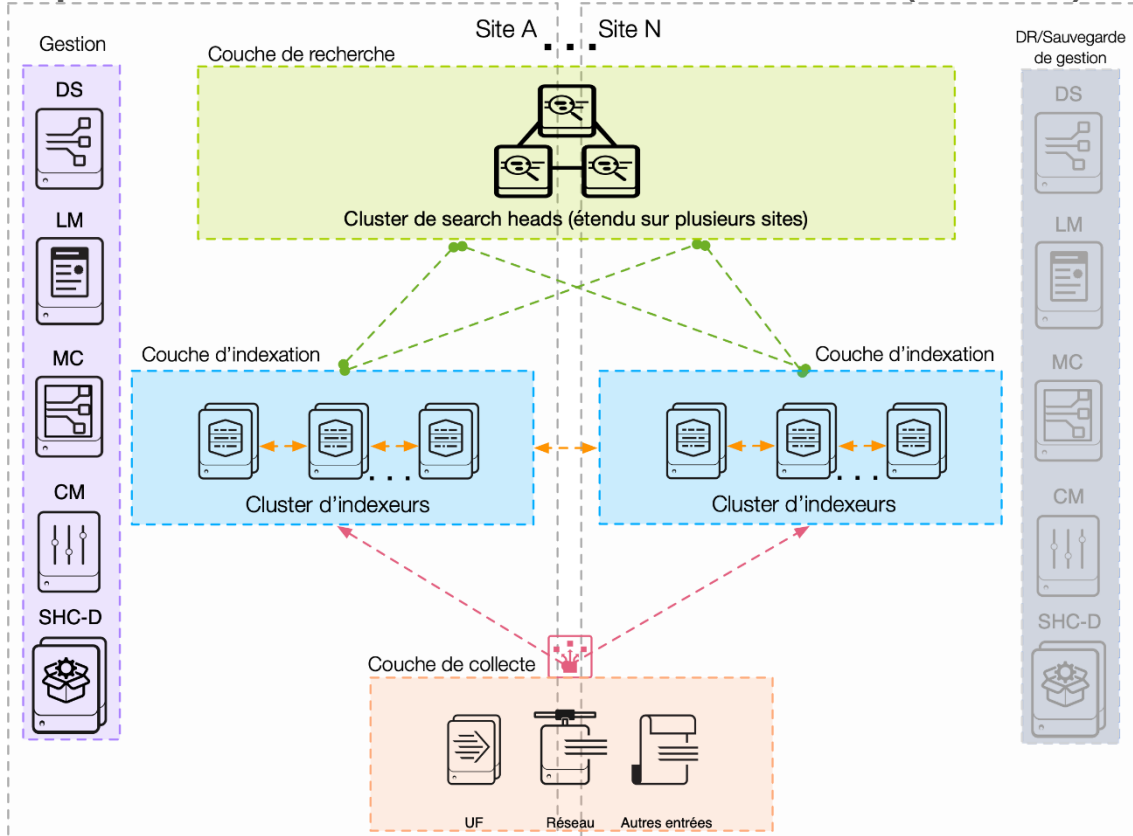
Déploiement en cluster distribué + SHC - Multisite (M3 / M13)



Description du déploiement en cluster distribué + SHC - Multisite (M3 / M13)	Limitations
<p>Cette topologie ajoute une évolutivité horizontale et élimine le point de défaillance individuel intervenant au niveau de la couche de recherche sur chaque site. Il faut au moins trois search heads (par site) pour déployer un SHC.</p> <p>Pour gérer la configuration du SHC, un autre composant Splunk, appelé Déployeur de cluster de search heads, est nécessaire pour chaque SHC. Ce composant est nécessaire pour déployer les modifications dans les fichiers de configuration du cluster. Le Déployeur de cluster de search heads n'implique aucune spécification de haute disponibilité (aucun rôle runtime).</p> <p>Le SHC apporte les avantages suivants : a) augmentation de la capacité en search heads au-delà de ce qu'une seule search head peut fournir, b) planification de la répartition de la charge de recherche dans le cluster, et c) basculement optimal des utilisateurs en cas de défaillance d'une search head.</p> <p>Un équilibreur de charge réseau prenant en charge les sessions persistantes doit également être présent devant chaque membre du SHC (et ce, sur chaque site), afin de garantir un équilibrage adéquat des utilisateurs dans l'ensemble du cluster.</p> <p>Remarque à l'attention des clients ES : si votre code de catégorie est M13 (autrement dit, si vous prévoyez de déployer l'application Splunk dédiée à la sécurité des entreprises), un seul cluster de search heads dédié présent sur <i>un site</i> est nécessaire au déploiement de l'application (non explicitement représenté dans le schéma de topologie). Pour pouvoir rétablir</p>	<ul style="list-style-type: none"> Aucune réplication des artefacts de recherche entre les sites, les SHC sont autonomes La latence inter-sites affectant la réplication des index doit être conforme aux limites documentées Le SHC ne peut pas comprendre plus de 100 nœuds.

Description du déploiement en cluster distribué + SHC - Multisite (M3 / M13)	Limitations
un environnement de SH ES après une défaillance de site, nous pouvons utiliser une technologie tierce pour faire basculer les instances de search heads ou provisionner une SH ES en « veille à chaud » et la maintenir synchronisée avec l'environnement de SH principal. Nous vous recommandons vivement de faire appel aux Services professionnels Splunk pour déployer ES dans un environnement HA/DR.	

Déploiement en cluster distribué + SHC - Multisite (M4 / M14)



Description du déploiement en cluster distribué + SHC - Multisite (M4 / M14)	Limitations
<p>Il s'agit de l'architecture validée la plus complexe. Elle est conçue pour les déploiements qui doivent répondre à des exigences strictes en matière de haute disponibilité et de récupération en cas de sinistre. Nous vous recommandons vivement de faire appel aux Services professionnels Splunk pour réaliser un déploiement approprié. Lorsqu'elle est correctement déployée, cette topologie assure le fonctionnement continu de votre infrastructure Splunk pour la collecte de données, leur indexation et la recherche.</p> <p>Cette topologie implique le déploiement d'un cluster de search heads « étendu » qui couvre un ou plusieurs sites. Vous obtenez ainsi un mécanisme de basculement optimal des utilisateurs en cas de défaillance d'un nœud de recherche ou d'un datacenter. Les artefacts de recherche et autres objets de connaissance runtime sont répliqués dans le SHC. Une configuration minutieuse est nécessaire pour faire en sorte</p>	<ul style="list-style-type: none"> La latence du réseau entre les sites doit être conforme aux limites documentées Le basculement du SHC peut nécessiter une intervention manuelle dans le cas où seule une minorité de clusters survit.

Description du déploiement en cluster distribué + SHC - Multisite (M4 / M14)	Limitations
<p>que la réplication s'exécute entre les sites, dans la mesure où le SHC n'a pas connaissance de son site (autrement dit, la réplication des artefacts n'est pas déterministe).</p> <p>Vous pouvez configurer l'affinité des sites de manière à ce que la liaison WAN entre les sites ne soit exploitée qu'en cas d'impossibilité d'exécuter une recherche localement.</p> <p>Un équilibreur de charge réseau prenant en charge les sessions persistantes doit également être présent devant chaque membre du SHC, afin de garantir un équilibrage adéquat des utilisateurs dans l'ensemble du cluster.</p> <p>Remarque à l'attention des clients ES : si votre code de catégorie est M14 (autrement dit, si vous prévoyez de déployer l'application Splunk dédiée à la sécurité des entreprises), un seul cluster de search heads dédié présent sur <i>un site</i> est nécessaire au déploiement de l'application (non explicitement représenté dans le schéma de topologie). ES requiert un ensemble cohérent d'artefacts de runtime pour être accessible, ce qui ne peut pas être garanti dans un SHC étendu en cas de panne d'un site. Pour pouvoir rétablir un environnement de SH ES après une défaillance de site, nous pouvons utiliser une technologie tierce pour faire basculer les instances de search heads ou provisionner une SH ES en « veille à chaud » et la maintenir synchronisée avec l'environnement de SH principal. Nous vous recommandons vivement de faire appel aux Services professionnels Splunk pour déployer ES dans un environnement HA/DR.</p>	

Étape 1b : définir vos exigences en collecte de données

La couche de collecte de données est un composant essentiel du déploiement Splunk. Elle permet à n'importe quelle machine de votre environnement de transmettre des données à la couche d'indexation à des fins de traitement, pour les rendre interrogeables dans Splunk. Cette fois, l'aspect le plus important réside dans l'assurance d'une efficacité et d'une fiabilité maximales de la transmission et de l'indexation, qui sont des facteurs essentiels pour les performances et la réussite de votre déploiement Splunk.

Tenez compte des aspects suivants lors de l'élaboration de l'architecture de la couche de collecte de données :

- Origine de vos données. Proviennent-elles de fichiers log, de sources syslog, d'entrées réseau, de systèmes de journalisation d'événements de système d'exploitation, d'applications, de bus de messages ou d'ailleurs ?
- Exigences en latence et en débit d'ingestion des données
- Répartition idéale des événements sur les indexeurs de votre couche d'indexation
- Tolérance aux défaillances et récupération automatique (haute disponibilité ou HA)
- Exigences de sécurité et de souveraineté des données

Cette section des AVS présente les méthodes courantes de collecte de données. Elle aborde également les architectures et les bonnes pratiques propres à chaque méthode de collecte tout en mettant en lumière les problèmes potentiels à envisager lors des choix de déploiement.

Considérations essentielles relatives à l'architecture et raisons de leur importance

Du fait du rôle crucial que joue la couche de collecte de données, il est important de comprendre les considérations clés impliquées dans la conception de l'architecture.

Bien que certaines de ces considérations ne vous concernent pas nécessairement, celles qui figurent en gras dans le tableau ci-dessous décrivent des éléments fondamentaux qui s'appliquent à tous les environnements.

Considération	Raison de son importance
Les données sont correctement ingérées (horodatage, fins de ligne, troncage)	Nous ne saurions trop insister sur l'importance d'une répartition idéale des événements entre les indexeurs. La couche d'indexation fonctionne au mieux lorsque tous les indexeurs disponibles sont équitablement exploités. C'est aussi vrai pour l'ingestion des données que pour les performances des recherches. Un indexeur qui assure un volume d'ingestion de données très supérieur à celui de ses pairs peut avoir un impact négatif sur les délais de réponse des recherches. Pour les indexeurs qui disposent d'un espace disque local limité, une répartition inégale des événements peut également entraîner une péremption prématurée des données qui serait en infraction avec la politique de conservation des données configurée.
Les données sont réparties de façon optimale sur tous les indexeurs disponibles	Si les données ne sont pas correctement ingérées car l'horodatage des événements et les fins de ligne ne sont pas bien configurés, il deviendra très difficile de les interroger. En effet, la délimitation des événements doit se faire au moment de la recherche. Si la configuration de l'extraction des horodatages est erronée ou manquante, un horodatage implicite indésirable risque d'être appliqué. Une telle situation sera source de confusion pour vos utilisateurs, et il sera beaucoup plus difficile que prévu d'extraire de la valeur de vos données.
Toutes les données atteignent la couche d'indexation de façon fiable et sans pertes	Toutes les données de log qui sont recueillies pour mener des analyses fiables doivent être exhaustives et valides, afin que les recherches qui s'y appliquent fournissent des résultats valables et précis.
Toutes les données atteignent la couche d'indexation avec une latence minimale	Les retards dans l'ingestion des données augmentent le délai entre la survenue d'un événement potentiellement critique et la possibilité de l'interroger et d'y réagir. Il est souvent crucial d'assurer une latence d'ingestion minimale dans les scénarios de surveillance qui déclenchent des alertes pour avertir du personnel ou déclencher des actions automatisées.
Les données sont sécurisées durant leur transit	Si les données sont sensibles ou doivent être protégées durant leur envoi sur des réseaux non contrôlés, il peut s'avérer nécessaire de les chiffrer pour éviter leur interception par des tiers non autorisés. Nous recommandons généralement d'activer le SSL pour toutes les connexions entre les composants Splunk.
L'utilisation des ressources réseau est réduite au minimum	L'impact sur les ressources réseau de la collecte des données de log doit être réduit au minimum pour éviter d'affecter les autres trafics réseau stratégiques. Dans le cas des réseaux à ligne louée, la réduction de l'utilisation du

	réseau contribue également à la diminution du TCO de votre déploiement.
Authentification/autorisation des sources de données	Pour éviter que des sources de données incontrôlables n'affectent votre environnement d'indexation, pensez à déployer un mécanisme d'authentification/d'autorisation des connexions. Pour ce faire, vous pouvez employer des contrôles réseau ou des mécanismes au niveau des applications (SSL/TLS par exemple).

En raison de son rôle vital pour votre déploiement, les consignes de ce document ciblent les architectures qui assurent une répartition idéale des événements. Lorsqu'un environnement Splunk n'offre pas les performances de recherche attendues, dans tous les cas ou presque, cela s'explique par des performances de stockage inférieures aux spécifications minimales et/ou par une répartition inégale des événements, qui limite l'exploitation de la parallélisation des recherches.

Maintenant que vous comprenez les aspects architecturaux les plus cruciaux, nous allons aborder les exigences qu'il faut respecter en matière de collecte de données.

Questionnaire 2 : définir vos exigences en collecte de données

En répondant aux questions suivantes, vous obtiendrez une liste des composants de collecte de données dont vous avez besoin dans votre déploiement. Le code figurant dans la colonne la plus à droite vous permettra de trouver des informations supplémentaires sur chaque composant dans la suite du document.

#	Question	Considérations	Impact sur la topologie	Composants de collecte de données concernés
1	Avez-vous besoin de surveiller des fichiers locaux ou d'exécuter des scripts de collecte de données sur des points de terminaison ?	Il s'agit d'une exigence fondamentale dans la plupart des scénarios de déploiement de Splunk.	Vous devrez installer des forwarders universels sur vos points de terminaison et gérer leur configuration de manière centralisée.	UF
2	Avez-vous besoin de recueillir des données de log envoyées via syslog par des périphériques sur lesquels vous ne pouvez pas installer de logiciels (dispositifs, commutateurs réseau, etc.) ?	Syslog est un protocole de transport commun souvent employé par les périphériques spécialisés qui ne permettent pas l'installation de logiciels spécialisés.	Vous aurez besoin d'une infrastructure de serveur syslog qui jouera le rôle de point de collecte.	SYSLOG HEC
3	Avez-vous besoin de prendre en charge la collecte de données de log à partir d'applications qui se connectent à une API plutôt que d'écrire sur des disques locaux ?	L'écriture dans des fichiers log se trouvant sur des points de terminaison implique une prévision de l'espace disque et une gestion de ces fichiers log (rotation, suppression, etc.). Certains clients veulent abandonner ce modèle	Vous devrez utiliser le Collecteur d'événements HTTP (HEC) Splunk ou une autre technologie servant de collecteur de logs.	HEC

		et journaliser directement dans Splunk à l'aide des bibliothèques de journalisation disponibles.		
4	Avez-vous besoin de collecter des données auprès d'un fournisseur de flux de données d'événement ?	Beaucoup d'entreprises ont adopté un modèle de hub d'événements, dans lequel une plateforme centralisée de diffusion des données (comme AWS Kinesis ou Kafka) assure le transport des messages entre les producteurs de données de log et leurs utilisateurs.	Vous devrez prévoir l'intégration du fournisseur de flux de données à Splunk.	KAFKA KINESIS HEC
5	Appliquez-vous des politiques de sécurité non négociables qui empêchent les producteurs de logs d'établir directement des connexions TCP à la couche d'indexation ?	Les topologies réseau comprennent parfois plusieurs zones réseau, séparées par des règles de pare-feu restrictives. Il devient alors impossible d'autoriser globalement le trafic des ports Splunk à circuler entre ces zones. Il serait très fastidieux de configurer et de tenir à jour des règles de pare-feu individuelles pour chaque adresse IP source ou cible.	Vous aurez besoin d'une couche intermédiaire de transmission qui permettra au trafic de circuler entre les zones réseau.	IF
6	Avez-vous besoin de recueillir des données de log par des moyens programmatiques, par exemple, en invoquant des API REST ou en interrogeant des bases de données ?	Splunk fournit différentes entrées modulaires qui permettent d'exécuter des scripts sur des API dans un large éventail de scénarios d'ingestion de données, dont DBX pour la collecte de données à partir de bases de données relationnelles.	Dans votre couche de collecte, il faudra qu'un ou plusieurs nœuds de collecte de données (DCN) soient équipés d'un forwarder lourd Splunk.	DCN
7	Avez-vous besoin d'acheminer des données (ou un sous-jeu de données) vers d'autres systèmes en dehors (et en plus) de Splunk ?	Dans certains scénarios d'utilisation, les données indexées dans Splunk doivent également être transmises à un autre système. Il est fréquent que les données transmises ne constituent qu'un sous-jeu des données source, ou que les données doivent être modifiées avant d'être transmises.	Selon les spécificités du scénario d'utilisation, vous aurez peut-être besoin d'une couche de transmission intermédiaire équipée d'un forwarder lourd pour prendre en charge l'acheminement et le filtrage reposant sur les événements. Vous pouvez également transmettre les données après leur indexation en	HF

			utilisant la commande cefout de l'application Splunk pour CEF.	
8	Votre environnement comprend-il des sites distants soumis à des contraintes de bande passante exigeant un filtrage significatif des données avant de les envoyer sur le réseau ?	Filtrer les données avant de les transmettre implique l'utilisation d'un forwarder capable de les analyser (forwarder lourd ou HF). La bande passante sortante utilisée par un HF est environ 5 fois supérieure à celle d'un forwarder universel, donc le filtrage n'a d'intérêt que s'il élimine un grand nombre d'événements (en général : plus de 50 % des données source). Idéalement, vous devrez adapter la granularité de la journalisation de manière à atteindre la réduction voulue du volume des logs.	Si vous ne pouvez pas réduire le volume des logs à la source, vous aurez besoin d'un HF intermédiaire sur le site distant, pour qu'il analyse les données source et filtre les événements en fonction de sa configuration.	IF HF
9	Avez-vous besoin de masquer/d'obscurcir des données sensibles avant qu'elles ne soient envoyées sur un réseau public à des fins d'indexation ?	La sécurisation du trafic des forwarders par SSL ne suffit parfois pas pour protéger les données sensibles durant leur transit sur des réseaux publics ; certaines parties des événements doivent alors être masquées avant transmission (numéro de sécurité sociale, données de carte bancaire, etc.). Idéalement, ce masquage doit être réalisé dans l'application qui produit les données de log.	Si vous ne pouvez pas masquer les données dans l'application qui les produit, vous aurez besoin d'un HF intermédiaire local qui analysera les données source et appliquera les règles de masquage voulues par la configuration, avant que les données ne soient envoyées vers les indexeurs.	IF HF
10	Avez-vous besoin de collecter des métriques à l'aide de statsd ou collectd ?	statsd et collectd sont des technologies communes utilisées pour recueillir des métriques auprès des systèmes hôtes et des applications.	Splunk prend en charge des types d'index et des méthodes de collecte spécifiques pour alimenter ces index à l'aide d'UF, de HF ou de HEC.	MÉTRIQUES
11	Avez-vous besoin d'une haute disponibilité pour certains composants de collecte de données ?	Bien qu'elle ne s'applique généralement pas aux points de terminaison, la disponibilité peut être un facteur clé pour d'autres composants de collecte de données, comme les forwarders intermédiaires ou les nœuds de collecte.	Il faudra donc tenir compte de l'impact que des pannes pourraient avoir sur la disponibilité de chaque composant afin de le minimiser.	HA

Étape 2b : choisir vos composants de collecte de données

Une fois que vous aurez rempli le questionnaire, vous obtiendrez une liste des composants de collecte de données nécessaires pour respecter les spécifications de votre déploiement. Cette section aborde les différents composants architecturaux de collecte de données de façon plus détaillée. Mais avant cela, quelques consignes générales.

Consignes générales pour l'architecture de transmission

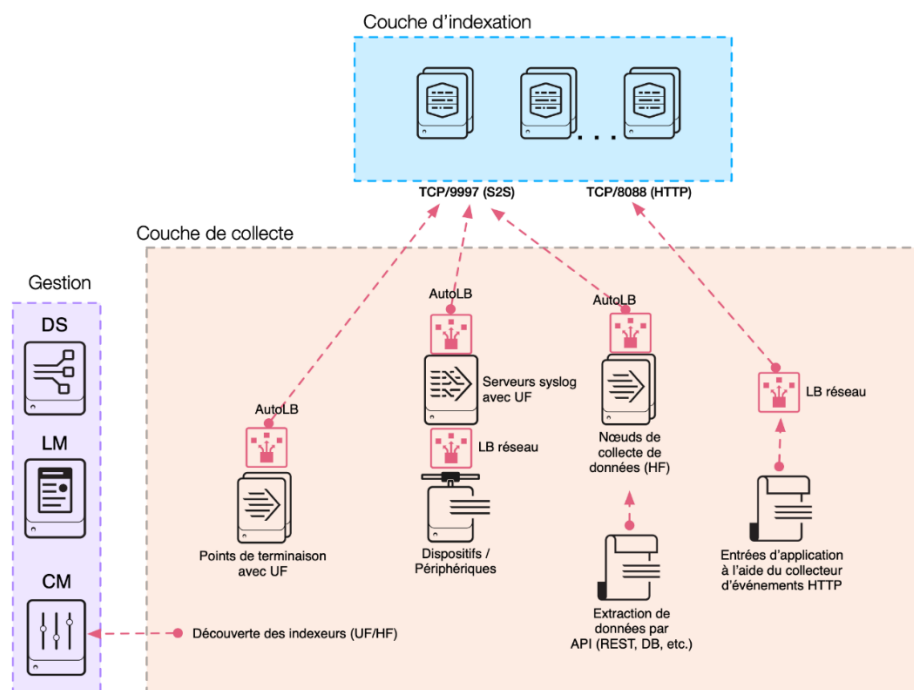
Idéalement, la couche de collecte de données doit être aussi « plate » que possible. Cela signifie que la collecte auprès des sources de données intervient localement par le biais d'un forwarder universel. Ces données sont ensuite transmises directement à la couche d'indexation. Cette approche est une bonne pratique car elle garantit une latence minimale de l'ingestion des données (délai avant recherche) et permet une répartition uniforme des événements sur les indexeurs disponibles. Elle a également pour effet de simplifier la gestion et l'exploitation des données. Nous rencontrons souvent des clients déployant une couche de transmission intermédiaire. En règle générale, il vaut mieux l'éviter à moins que vos exigences ne puissent être satisfaites autrement. En raison de l'impact potentiel des forwarders intermédiaires, une section distincte de ce document est consacrée à ce sujet.

Certains points de terminaison ne permettent pas l'installation d'un forwarder universel (autrement dit, les périphériques réseau, les dispositifs, etc.) et utilisent le protocole syslog pour leur journalisation. Une architecture recommandée distincte, permettant de collecter des données de ces sources de données, est décrite dans la section intitulée Collecte des données Syslog.

Dans le cas des sources pour lesquelles la collecte doit être effectuée à l'aide de moyens programmatiques (API, accès à une base de données), nous vous recommandons de déployer un nœud de collecte de données (DCN) basé sur une installation complète de Splunk Enterprise. C'est ce que l'on appelle un forwarder lourd (HF). Nous vous recommandons de ne pas exécuter ces types d'entrée sur la couche de search heads ailleurs que dans un environnement de développement.

Le schéma suivant illustre une architecture générale de collecte de données qui respecte cette consigne.

Présentation des topologies de collecte de données



Le schéma ci-dessus illustre un serveur de déploiement (DS) dans la couche de gestion utilisée pour gérer les configurations des composants de collecte. Le maître de licences (LM) apparaît également car les nœuds de collecte ont besoin d'y accéder pour utiliser les fonctionnalités de Splunk Enterprise. Le maître de clusters (CM), s'il est disponible, peut être utilisé par les forwarders à des fins de découverte d'indexeurs, ce qui évite d'avoir à gérer les indexeurs disponibles dans la configuration de sortie des forwarders.

Dans le schéma ci-dessus, AutoLB représente le mécanisme intégré d'équilibrage automatique de charge de Splunk. Ce mécanisme est utilisé pour garantir une répartition uniforme des événements pour les données envoyées à l'aide du protocole propriétaire S2S de Splunk (port 9997 par défaut). Remarque : l'utilisation d'un équilibrage des charges réseau externe pour le trafic S2S n'est actuellement pas pris en charge et il n'est pas recommandé.

Pour équilibrer la charge du trafic de sources de données qui communiquent à l'aide d'un protocole standard (comme le protocole HTTP ou syslog), nous utilisons un mécanisme d'équilibrage des charges pour assurer une répartition uniforme des charges et des événements sur les indexeurs de la couche d'indexation.

Forwarder universel (UF)

Un forwarder universel (UF) est le meilleur choix pour un large éventail d'exigences de collecte auprès des systèmes de votre environnement. Il s'agit d'un mécanisme de collecte spécialisé qui utilise très peu de ressources. L'UF doit être l'option par défaut pour la collecte et la transmission des données de log. L'UF fournit :

- une fonction de point de contrôle et de redémarrage pour une collecte de données sans perte ;
- un protocole efficace qui minimise l'utilisation de la bande passante réseau ;
- des capacités de réduction des événements ;
- un mécanisme intégré d'équilibrage des charges sur tous les indexeurs disponibles ;
- un chiffrement réseau optionnel par SSL/TLS ;
- une compression des données (uniquement utilisable sans SSL/TLS) ;
- plusieurs méthodes d'entrée (fichiers, journaux d'événements Windows, entrées réseau, entrées scriptées) ;
- des capacités limitées de filtrage des événements (journaux d'événements Windows uniquement) ;
- la prise en charge d'un pipeline d'ingestion des données pour augmenter le débit/réduire la latence.

À quelques exceptions près concernant les données bien structurées (json, csv, tsv), l'UF ne convertit pas les sources de log en événements et ne peut donc pas exécuter d'actions nécessitant une compréhension du format des logs. Il est également fourni avec une version allégée du langage Python, ce qui le rend incompatible avec l'ensemble des applications d'entrée modulaire qui nécessitent une pile Python complète pour fonctionner.

Il est courant de déployer un grand nombre d'UF (100 à 10 000) sur les points de terminaison et les serveurs d'un environnement Splunk, et de les gérer de manière centralisée, que ce soit via un serveur de déploiement Splunk ou un outil tiers de gestion des configurations (comme Puppet ou Chef).

Forwarder lourd (HF)

Le forwarder lourd (HF) est un déploiement complet de Splunk Enterprise configuré pour agir comme forwarder, sans aucune indexation. Le HF n'exécute généralement aucune autre fonction de Splunk. La différence essentielle entre un UF et un HF réside dans le fait que le HF contient l'ensemble du pipeline d'analyse et exécute toutes les fonctions d'un indexeur sans écrire les événements sur un disque ou les indexer. Cela permet au HF de comprendre les différents événements et d'agir en conséquence, par exemple, pour masquer des données ou

appliquer un filtrage et un routage en fonction des données d'événement. Étant donné qu'il comprend une installation complète de Splunk Enterprise, il peut héberger des entrées modulaires qui nécessitent l'ensemble de la pile Python pour assurer correctement les fonctions de collecte de données ou pour servir de point de terminaison au collecteur d'événements HTTP (HEC) de Splunk. Le HF comprend les fonctions et caractéristiques suivantes :

- analyse des données en tant qu'événements ;
- filtrage et acheminement des événements en fonction des données d'événement ;
- utilisation de davantage de ressources que l'UF ;
- utilisation de davantage de bande passante réseau que l'UF (env. 5 fois plus) ;
- interface graphique de gestion.

En règle générale, les HF ne sont pas installés sur des points de terminaison pour collecter les données. Ils sont plutôt utilisés sur des systèmes autonomes pour implémenter des nœuds de collecte (DCN) ou des couches de transmission intermédiaires. **Réservez l'utilisation des HF aux seuls cas dans lesquels la collecte des données d'un autre système ne peut pas s'effectuer avec un UF.**

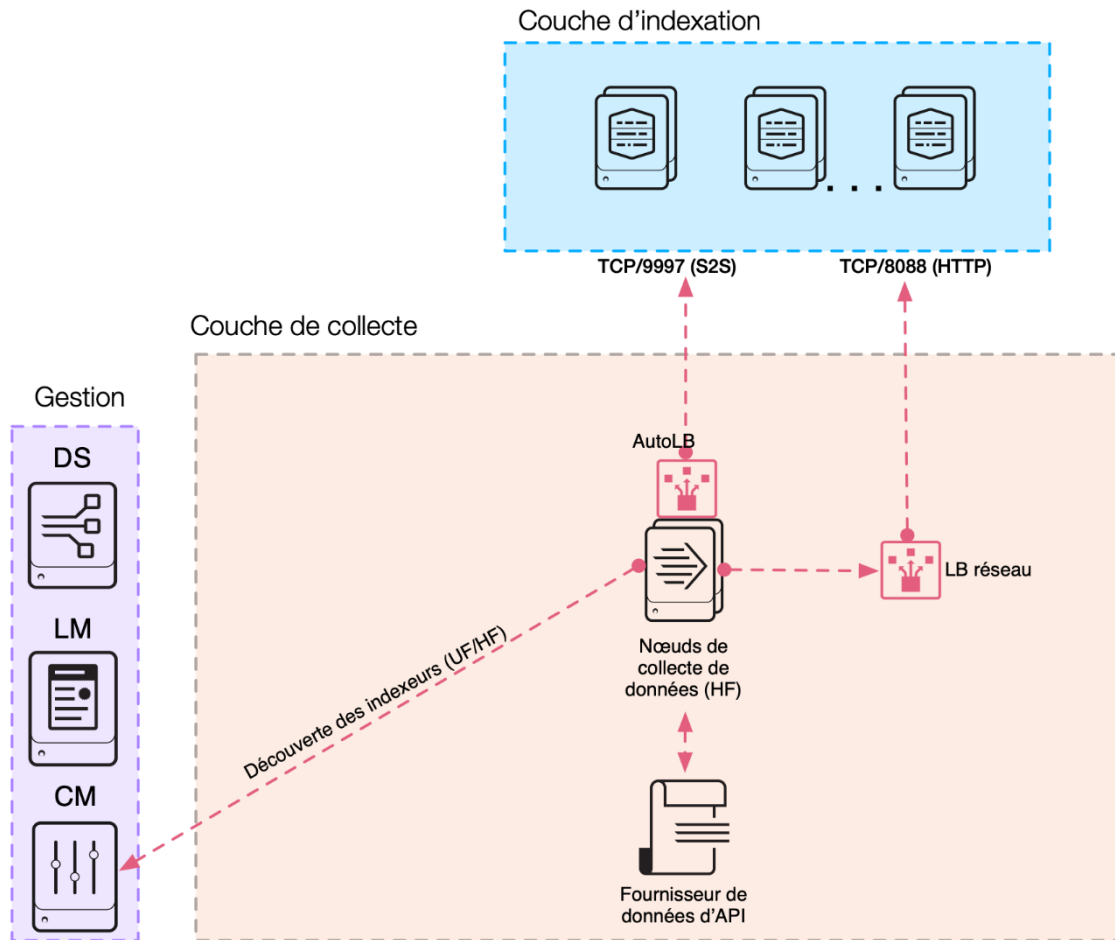
Quelques exemples de ces cas :

- lecture des données d'un RDBMS pour les ingérer dans Splunk (entrées de base de données) ;
- collecte de données auprès de systèmes accessibles via une API (services cloud, surveillance de VMWare, systèmes propriétaires, etc.) ;
- mise à disposition d'une couche dédiée pour héberger le service de collecteur d'événements HTTP ;
- mise en place d'une couche de transmission intermédiaire nécessitant un forwarder capable d'analyser à des fins de routage, de filtrage ou de masquage.

Forwarder lourd comme nœud de collecte de données (DCN)

Pour certaines sources de données, la collecte doit s'effectuer à l'aide d'une API quelconque. Les mécanismes de requête peuvent varier : REST, services web, JMS et/ou JDBC. Splunk et des développeurs tiers proposent un large éventail d'applications qui permettent à ces interactions d'API d'avoir lieu. La plupart du temps, ces applications sont implémentées à l'aide du framework d'entrée modulaire, qui nécessite une installation de Splunk Enterprise pour fonctionner correctement. Pour ce scénario d'utilisation, la bonne pratique consiste à déployer un ou plusieurs serveurs qui agiront comme forwarders lourds configurés pour fonctionner en tant que nœuds de collecte.

Topologie des nœuds de collecte de données

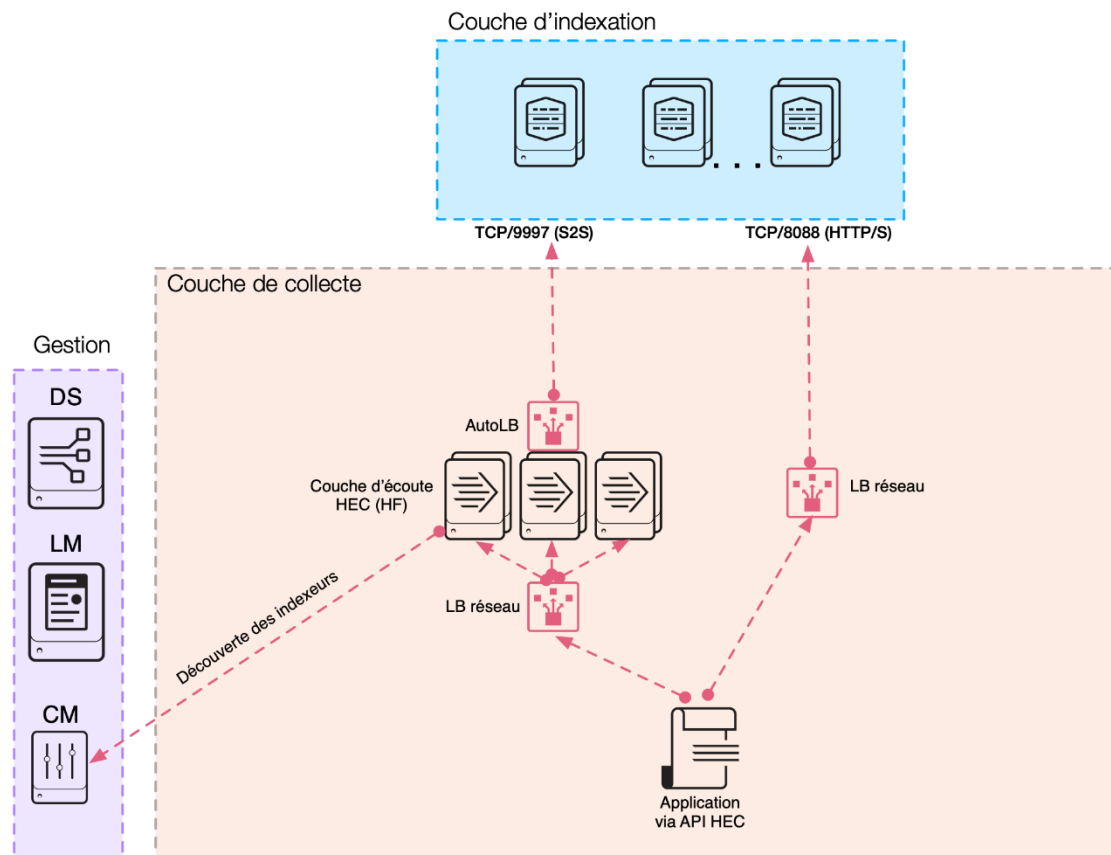


Collecteur d'événements HTTP (HEC)

Le HEC fournit un service d'écoute qui accepte les connexions HTTP/S côté serveur et une API côté client, ce qui permet aux applications d'envoyer des paquets de données de log directement vers la couche d'indexation ou vers une couche de réception HEC dédiée comprenant un ou plusieurs forwarders lourds. Le HEC fournit deux points de terminaison qui prennent en charge l'envoi de données au format brut ou JSON. L'utilisation du format JSON permet d'inclure des métadonnées dans le paquet d'événement pour plus de flexibilité lors des recherches qui porteront ultérieurement sur ces données.

Le schéma suivant illustre les deux options de déploiement de HEC :

Options de topologie pour HEC



La couche de gestion comprend le maître de licences (nécessaire au HF) ainsi que le serveur de déploiement qui gère les entrées HTTP sur les composants d'écoute. Remarque : si la couche d'indexation est en cluster et qu'elle reçoit directement le trafic HEC, la configuration du HEC est gérée via le maître de clusters et non par le serveur de déploiement.

Le choix de la topologie dépend largement de vos besoins spécifiques. Une couche d'écoute HEC dédiée introduit un autre composant architectural dans votre déploiement. L'avantage de cette approche est que cette couche peut évoluer de manière indépendante et fournir un certain degré d'isolation par rapport à la couche d'indexation du point de vue gestion. De plus, étant donné que la couche HEC dédiée nécessite un HF, celui-ci analysera l'ensemble du trafic entrant et déchargera ainsi les indexeurs de cette mission.

Par contre, héberger l'écoute HEC directement sur les indexeurs tend à assurer une meilleure répartition des événements dans la couche d'indexation, car le protocole HTTP est un protocole bien compris par tous les dispositifs d'équilibrage des charges réseau et car une politique d'équilibrage appropriée peut faire en sorte que les indexeurs les moins chargés soient alimentés en premier.

Dans l'optique de déployer l'architecture la plus simple répondant à vos exigences, nous vous recommandons d'envisager d'héberger l'écoute HEC sur les indexeurs, sous réserve que vous disposiez de capacités système suffisantes pour cela. Cette décision peut aisément être révoquée par la suite si le besoin s'en fait ressentir, tout simplement en déployant une couche HF de taille adaptée correctement configurée, et en modifiant la configuration de l'équilibrage des charges (LB) de façon à utiliser les adresses IP des HF plutôt que celle des indexeurs. Ce changement doit être transparent pour les applications clientes.

Remarque : si vous avez besoin que les indexeurs accusent réception des données envoyées via le HEC, nous vous recommandons d'employer une couche d'écoute HEC dédiée pour minimiser les messages en double causés par le redémarrage roulant des indexeurs.

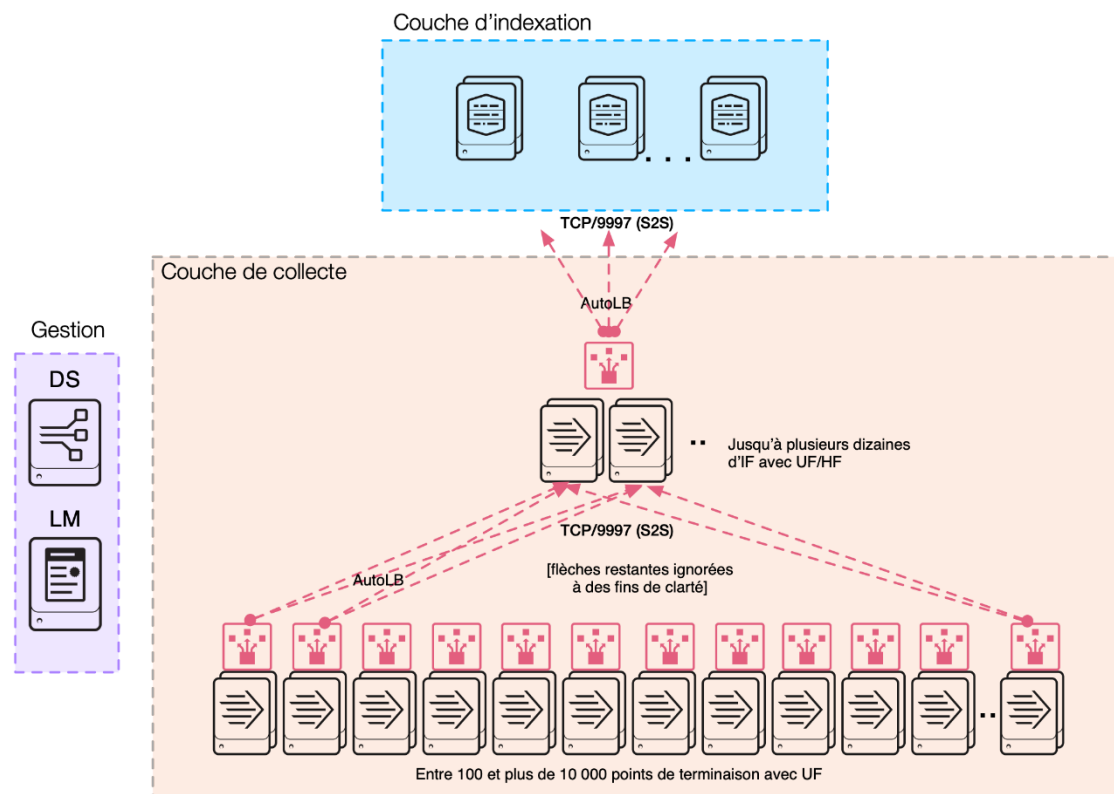
Remarque : cette architecture de déploiement de HEC assure le transport de certains autres éléments de collecte abordés plus loin, en particulier Syslog et la collecte de données de métriques.

Couche de transmission intermédiaire (IF)

Dans certaines situations, vous aurez recours à des forwarders intermédiaires pour transmettre les données. Les forwarders intermédiaires reçoivent les flux de log des points de terminaison et les transmettent à une couche d'indexeurs. Les forwarders intermédiaires introduisent des défis architecturaux qui exigent une conception soignée pour éviter tout impact négatif sur la globalité de l'environnement Splunk. En particulier, les forwarders intermédiaires concentrent les connexions de centaines, voire de milliers de forwarders de points de terminaison tout en transmettant les données aux indexeurs en utilisant un nombre de connexions beaucoup plus réduit. Cela peut considérablement affecter la répartition des données dans la couche d'indexation, car seule une partie des indexeurs reçoit du trafic à un moment donné. Ces inconvénients peuvent toutefois être compensés par de bons choix de dimensionnement et de configuration.

Le schéma suivant illustre bien ce problème :

Topologie de transmission intermédiaire



Dans un scénario comprenant un seul forwarder intermédiaire, tous les points de terminaison (potentiellement des milliers) se connectent à ce forwarder et le forwarder intermédiaire se connecte à son tour à un seul indexeur à la fois. Ce scénario n'est pas optimal car vous risquez d'être confronté aux conséquences suivantes :

- Un flux de données volumineux provenant de nombreux points de terminaison se déverse dans un seul et même canal qui épuise les ressources système et réseau.
- En cas de défaillance de l'IF, les destinations de basculement des points de terminaison sont limitées (le risque de panne est inversement proportionnel au nombre d'IF).

- Un petit nombre d'indexeurs peut être desservi simultanément. Les recherches portant sur de courtes périodes ne bénéficieront pas autant de la parallélisation qu'elles le pourraient autrement.

Les forwarders intermédiaires ajoutent également une couche d'architecture supplémentaire à votre déploiement, ce qui peut compliquer la gestion et le dépannage, et ajouter de la latence à votre parcours d'ingestion des données. Essayez d'éviter de recourir à une couche de transmission intermédiaire, excepté si c'est la seule façon de répondre à vos exigences. Vous pouvez envisager d'utiliser une couche intermédiaire si votre environnement présente les caractéristiques suivantes :

- Des données sensibles doivent être obscurcies ou supprimées avant transmission aux indexeurs du réseau. C'est notamment le cas si vous devez utiliser un réseau public.
- Des politiques de sécurité strictes n'autorisent pas les connexions directes entre les points de terminaison et les indexeurs, comme c'est le cas avec les réseaux multizones ou les indexeurs cloud.
- Des contraintes de bande passante entre les points de terminaison et les indexeurs imposent le filtrage d'une grande partie des événements.
- Les spécifications imposent un routage basé sur les événements vers des cibles dynamiques.

Réfléchissez attentivement au dimensionnement et à la configuration de l'ensemble de la couche de transmission intermédiaire pour assurer sa disponibilité. Fournissez une capacité de calcul suffisante pour gérer l'ensemble du trafic et assurez une répartition uniforme des événements sur les indexeurs. La couche IF implique les exigences suivantes :

- nombre total suffisant de pipelines de traitement des données ;
- infrastructure IF redondante ;
- configuration d'équilibrage des charges Splunk correctement ajustée. Par exemple, `autoLBVolume`, `EVENT_BREAKER`, `EVENT_BREAKER_ENABLE`, et éventuellement `forceTimeBasedAutoLB` si nécessaire.

La règle générale suggère de prévoir deux fois plus de pipelines de traitement IF que d'indexeurs de couche d'indexation.

Remarque : un pipeline de traitement n'est pas équivalent à un serveur IF physique. Avec des ressources système suffisantes, si des cœurs de processeur, de la mémoire et de la bande passante de carte réseau sont disponibles par exemple, un seul IF peut être configuré avec plusieurs pipelines de traitement.

Si vous avez besoin d'une couche IF ([voir le questionnaire](#)), optez par défaut pour des UF car ils offrent un débit supérieur tout en utilisant moins de ressources système et réseau. Utilisez un HF uniquement si les capacités des UF ne respectent pas vos spécifications.

Collecte des données Syslog (SYSLOG)

Le protocole syslog offre une source commune pour les données de log de toute l'entreprise. Les couches de collecte les plus évolutives et les plus fiables comprennent un composant d'ingestion de données syslog. Vous pouvez ingérer les données syslog dans Splunk de plusieurs façons. Étudiez les méthodes suivantes :

- **Forwarder universel (UF) / Forwarder lourd (HF) :** Utilisez un UF ou un HF Splunk pour surveiller (ingérer) les fichiers écrits par un serveur syslog (tel que `rsyslog` ou `syslog-ng`).
- **Agent Syslog vers HEC :** Utilisez un agent syslog capable de transmettre au HEC de Splunk. (Il existe des modules tiers pour `rsyslog` et `syslog-ng` qui disposent de cette capacité.).
- **Entrée TCP/UDP directe :** Splunk a la capacité d'écouter un port TCP ou UDP (port UDP 514 par défaut) et d'ingérer les données des sources de données (**non** recommandé en production).

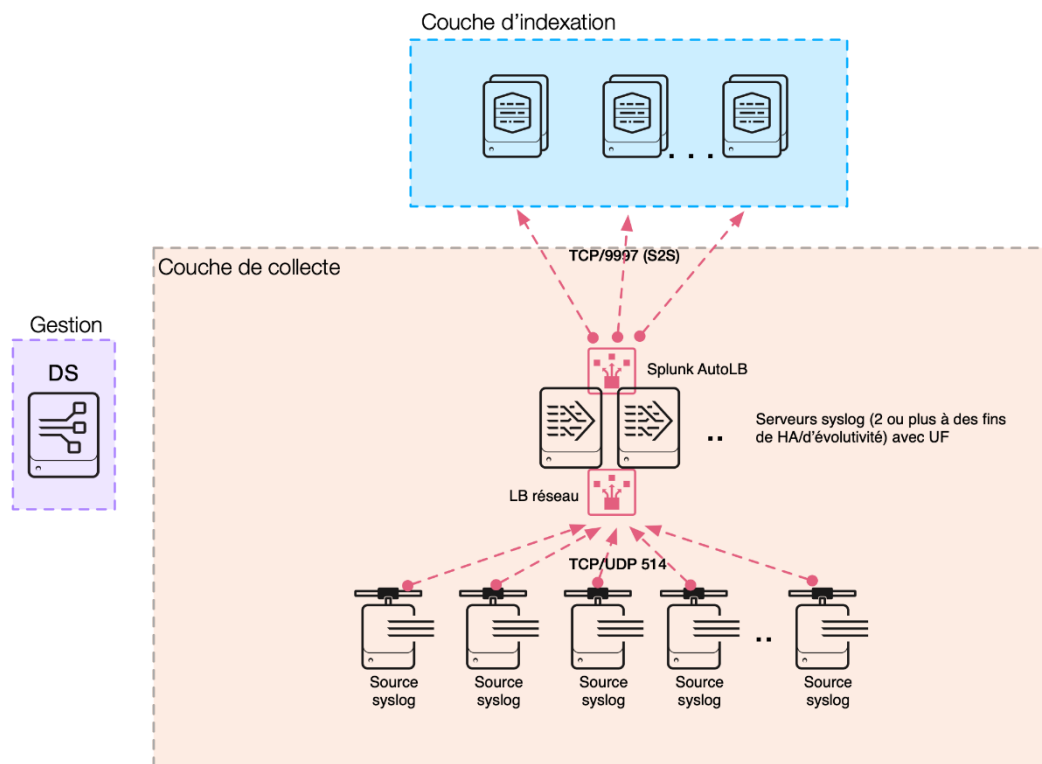
Syslog (surveillance des fichiers conjointement à un SCD)

Splunk peut, avec `inputs.conf`, surveiller un UF/HF pour traiter et ingérer les données des sources syslog qui sont écrites sur le disque au niveau d'un point de terminaison par un démon de collecte syslog (SCD). Les démons les plus courants, `rsyslog`, `syslog-ng` et [Fastvue](#), proposent des solutions commerciales et gratuites qui sont à la fois évolutives et simples à intégrer et à gérer, aussi bien dans les environnements à faible volume que dans les grands environnements distribués.

Pour en savoir plus sur la configuration des moniteurs, consultez [Surveillance des fichiers et des répertoires](#) dans *Injection des données*.

Cette architecture assure une bonne intégration des données, tout comme un forwarder universel le fait sur n'importe quel autre point de terminaison. Vous pouvez configurer le SCD de façon à identifier différents types de log et écrire les événements de log dans des fichiers et répertoires appropriés, où un forwarder Splunk pourra ensuite les prélever. Cela ajoute également un degré de persistance au flux syslog en écrivant les événements sur un disque, ce qui peut limiter le risque de perte de données pour les messages transportés par le protocole UDP, qui est peu fiable.

Topologie de collecte des données syslog à l'aide d'UF



Le schéma représente des sources syslog qui envoient des données à l'aide du protocole TCP ou UDP sur le port 514 vers un pool de serveurs syslog avec équilibrage des charges. Plusieurs serveurs garantissent la haute disponibilité de la couche de collecte et peuvent éviter les pertes de données durant les opérations de maintenance. Chaque serveur syslog est configuré pour appliquer des règles au flux syslog, ce qui permet d'écrire les événements syslog dans des fichiers/répertoires spécifiques à chaque type de source (événements de pare-feu, syslog de système d'exploitation, commutateurs réseau, IPS, etc.). L'UF qui est déployé sur chaque serveur surveille ces fichiers et transmet les données à la couche d'indexation pour qu'elles soient traitées dans l'index approprié. Splunk AutoLB se charge de répartir uniformément les données sur les indexeurs disponibles.

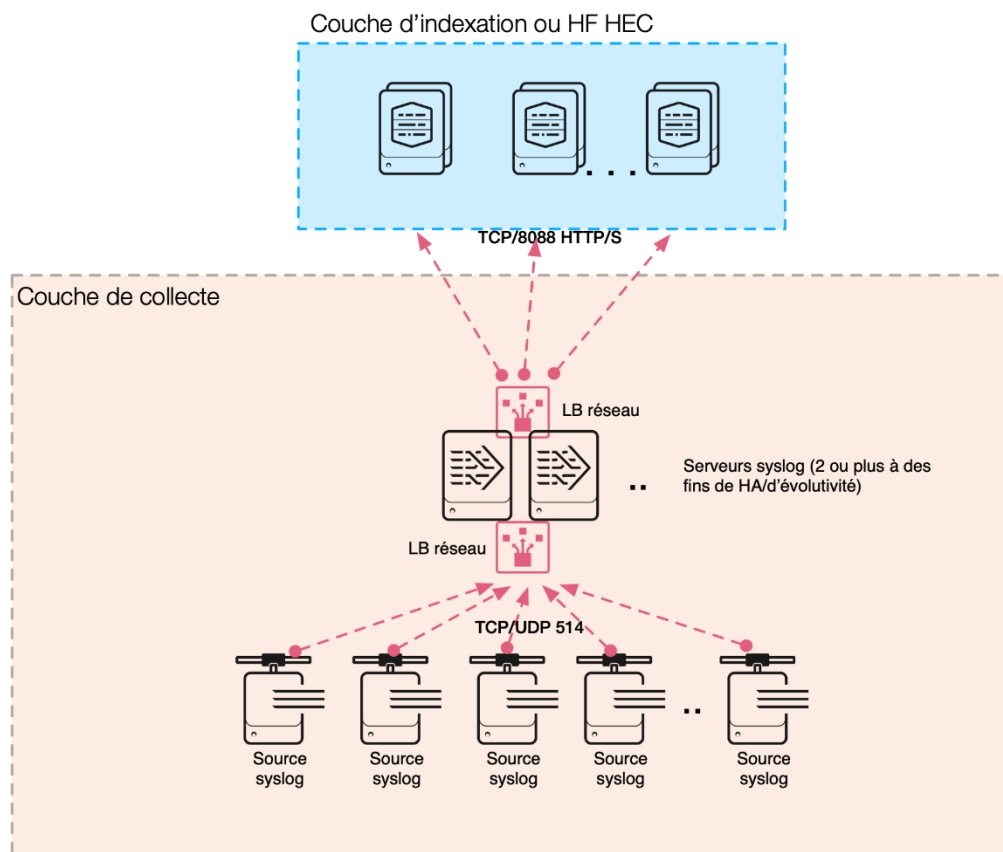
Le serveur de déploiement figurant dans la couche de gestion peut être utilisé pour gérer la configuration des UF de manière centralisée.

Agent Syslog vers HEC

Avec l'adoption de plus en plus répandue du HEC, un nombre croissant de déploiements l'utilisent pour ingérer les données syslog. Pour en savoir plus, consultez l'article [Syslog-ng et HEC : une collecte des données agrégées évolutive dans Splunk](#) sur le blog de Splunk.

Le schéma ci-dessous représente des sources syslog qui envoient des données sur le port 514 vers une ferme de serveurs syslog en utilisant un équilibrage des charges réseau. Des politiques syslog adaptées comprenant une destination personnalisée pour les données syslog (un script Python utilisant l'API HEC) sont appliquées et les événements sont envoyés à un écouteur HEC pour indexation, toujours avec un équilibrage des charges :

Topologie de collecte des données syslog à l'aide du HEC



Un avantage de cette topologie est qu'il n'est pas nécessaire de déployer et de configurer d'UF ou de HF. L'équilibrage des charges HTTP dessert les écouteurs HEC sur les indexeurs (ou une couche d'écoute HEC dédiée) pour assurer la répartition uniforme des données sur les points de terminaison HEC. Configurez cet équilibrage des charges en appliquant la politique « Le moins de connexions ».

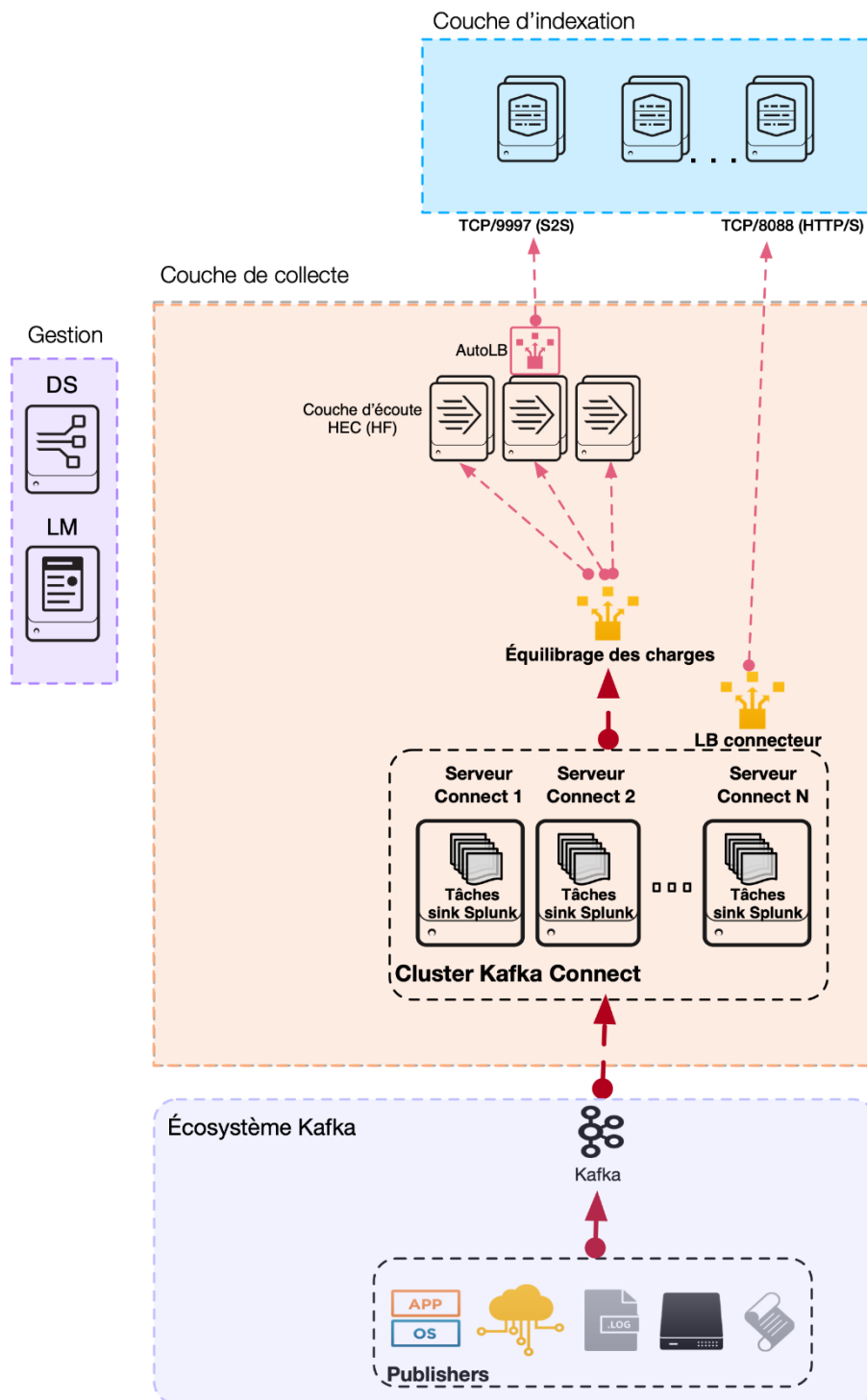
Entrée UDP Splunk

Splunk peut utiliser une entrée UDP directe sur un UF ou un HF pour recevoir des données syslog. Pour en savoir plus sur la configuration des ports TCP et UDP, consultez [Assimilation des données de ports TCP et UDP](#) dans *Injection des données*. La possibilité de recevoir des événements sur le port UDP 514 repose sur la capacité de l'UF ou du HF à s'exécuter en tant que racine. De plus, l'agent doit être disponible 100 % du temps pour éviter tout risque de perte de données. Les forwarders peuvent être fréquemment redémarrés pour appliquer des changements de configuration, ce qui est quasiment synonyme de perte de données. Pour ces raisons, **cette approche n'est pas considérée comme une bonne pratique pour un déploiement de production.**

Ingestion de données de log de rubriques Kafka (KAFKA)

Splunk fournit un connecteur de collecteur pour ingérer les données de rubriques Kafka : il s'agit de « Splunk Connect for Kafka ». Consultez [Apache Kafka Connect](#) dans le manuel de Splunk Connect for Kafka pour obtenir la documentation. Le package Splunk Connect for Kafka est installé sur un cluster Kafka Connect d'une taille adaptée (hors de Splunk), où il peut s'abonner à des rubriques selon sa configuration et envoyer les événements ingérés pour qu'ils soient indexés, en utilisant le HEC :

Topologie de collecte des données avec Kafka et HEC

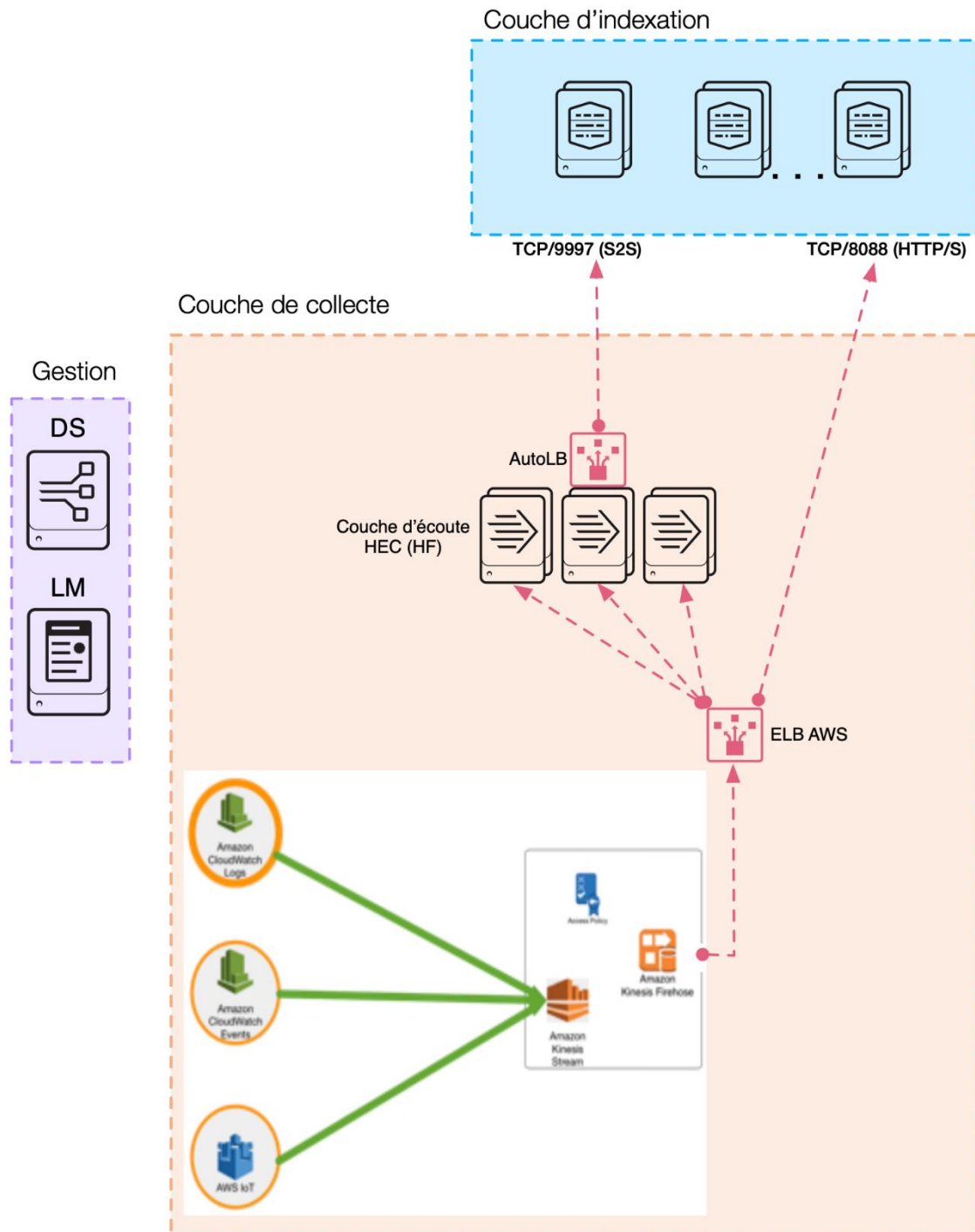


Le schéma représente les Publishers Kafka qui envoient des messages au bus Kafka. Les tâches hébergées dans le cluster Kafka Connect assimilent ces messages via Splunk Connect for Kafka et envoient les données au service d'écoute du HEC en utilisant un équilibrage des charges réseau. Une fois encore, le service d'écoute HEC peut être hébergé directement sur les indexeurs ou dans une couche d'écoute HEC dédiée. Veuillez consulter la section HEC pour plus d'informations. Les composants de la couche de gestion ne sont nécessaires que si une couche HF dédiée est déployée pour accueillir les écouteurs HEC.

Ingestion des données de log à partir d'Amazon Kinesis Firehose (KINESIS)

Splunk et Amazon ont implémenté une intégration entre Kinesis et le HEC Splunk, ce qui vous permet de diffuser des données d'AWS directement vers un point de terminaison HEC. Cette intégration est configurable dans la console AWS. Elle est complétée par l'[Extension Splunk pour Kinesis Firehose](#) qui fournit des connaissances CIM concernant différentes sources de données originaires d'AWS.

Topologie de collecte de données avec Amazon Kinesis



Le schéma représente des données de sources de log AWS envoyées à l'aide d'un flux Kinesis au Firehose qui, avec une configuration adaptée, enverra à son tour les données au service d'écoute HEC via un ELB AWS. Une fois encore, le service d'écoute HEC peut être hébergé directement sur les indexeurs ou dans une couche d'écoute HEC dédiée. Veuillez consulter la section HEC pour plus d'informations.

Les composants de la couche de gestion représentés ne sont nécessaires que si une couche HF dédiée est déployée pour accueillir les écouteurs HEC.

Collecte des métriques (MÉTRIQUES)

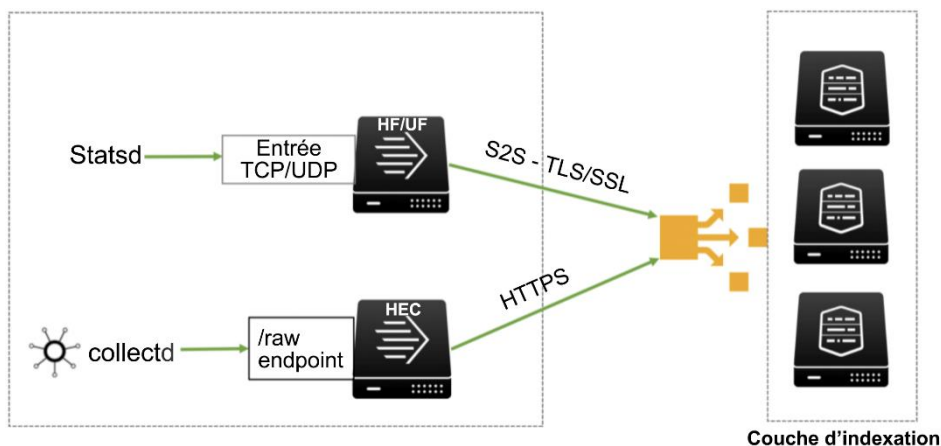
Splunk a la capacité de recevoir et de collecter des données de performances (ou métriques) d'un large éventail de systèmes et applications tiers. Les métriques de la plateforme Splunk utilisent un type d'index personnalisé qui est optimisé pour le stockage et la récupération des métriques.

Il existe différentes manières d'ingérer les données de métriques et la méthode de collecte dépend de la technologie employée. Le plus souvent, la collecte des métriques s'effectue au moyen d'un démon comme **collectd** ou **statsd**, ou d'un fichier de données de métriques personnalisé accompagné d'une configuration valide pour la source de données.

Il existe principalement deux méthodes pour ingérer des données dans Splunk lorsque l'on utilise des agents comme **statsd** et **collectd** : en utilisant une entrée directe **TCP/UDP** ou via le **HEC**.

L'utilisation du **HEC** est considérée comme une bonne pratique en raison de la résilience et de l'évolutivité du point de terminaison **HEC**, et parce qu'il est facile d'étendre horizontalement la couche de collecte.

Topologie de collecte des données de métriques



Statsd prend actuellement en charge le transport par **UDP** ou **TCP**, que vous pouvez utiliser comme entrée directe sur un forwarder Splunk ou un indexeur. Il n'est toutefois pas indiqué d'envoyer directement le trafic TCP/UDP aux forwarders en production, car l'architecture n'est pas résiliente et les pertes d'événements sont courantes (voir Collecte de syslog), en raison des redémarrages nécessaires des forwarders Splunk.

Considérations relatives à la haute disponibilité (HA) pour les composants de la couche de transmission

Le concept de haute disponibilité (HA) est omniprésent dans le monde numérique. Toutefois, d'une entreprise à l'autre, son sens peut varier pour être plus proche de la récupération en cas de sinistre (DR) que de la haute disponibilité à proprement parler. Bien qu'ils aient des points communs, ces deux concepts restent distincts. La HA caractérise un système qui vise à assurer un niveau convenu de performances opérationnelles (généralement son temps d'activité) pendant une période supérieure à la normale. La DR implique un ensemble de politiques, d'outils et de procédures qui permettent la récupération ou le maintien d'infrastructures technologiques et de systèmes vitaux à la suite d'une catastrophe.

La section suivante décrit les différentes formes de HA au niveau de la couche intermédiaire/d'agrégation :

Couche intermédiaire

- Pour les clients ayant déployé une couche intermédiaire/d'agrégation, la HA des forwarders est cruciale. Au niveau de la couche Applications, Splunk ne propose pas nativement de méthode de HA. Il existe d'autres stratégies pour assurer la HA au niveau du système d'exploitation, en dehors de Splunk. Les solutions les plus courantes sont VMWare VMotion, AWS Autoscaling Groups et Linux Clustering. Consultez votre Architecte Splunk pour discuter de ces options de conception.
- Dans les environnements qui imposent une exigence de HA à la couche HEC dédiée, la bonne pratique consiste à utiliser un équilibrage des charges de trafic réseau (NTLB) comme NGINX devant plusieurs forwarders lourds Splunk. Vous bénéficiez ainsi d'un débit, d'une échelle et d'une disponibilité maximisés. Nous obtenons un pool dédié d'instances de collecteur d'événements HTTP dont la seule mission consiste à recevoir et transmettre les données. Vous pouvez ajouter d'autres instances de HEC sans avoir à ajouter de nouveaux indexeurs. Si les indexeurs deviennent un goulot d'étranglement, ajoutez-en d'autres.
- Dans les environnements qui imposent une exigence de HA à la collecte syslog, la bonne pratique consiste à utiliser plusieurs serveurs syslog desservis par une adresse IP (virtuelle) de cluster, hébergée par une solution d'équilibrage des charges telle que HAProxy ou F5, pour assurer un débit, une échelle et une disponibilité maximisés. Nous obtenons un pool dédié d'instances de Splunk dont la seule mission consiste à recevoir et transmettre les données. Vous pouvez ajouter d'autres instances sans avoir à ajouter de nouveaux indexeurs. Si les indexeurs deviennent un goulot d'étranglement, ajoutez-en d'autres.

Couche de transmission

- Au niveau de la couche de transmission (point de terminaison), la HA de l'agent dépend du système d'exploitation sous-jacent. Au strict minimum, il faut faire en sorte que tous les services qui implémentent la fonctionnalité de transmission redémarrent automatiquement lorsque le système d'exploitation hôte redémarre. En-dehors de cela, les bonnes pratiques applicables aux forwarders indiquent qu'il faut configurer et utiliser correctement AutoLB entre les forwarders et les différents indexeurs. Il faut également que les indexeurs accusent réception des données pour confirmer leur arrivée dans la couche d'indexation.

Étape 3 : appliquer les principes de conception et les bonnes pratiques

Vous trouverez ci-dessous les principes de conception et bonnes pratiques propres à chaque couche de déploiement.

Couches de déploiement

Les principes de conception des AVS couvrent l'ensemble des couches de déploiement suivantes :

Couche	Définition
Recherche	<ul style="list-style-type: none"> • Search heads
Indexation	<ul style="list-style-type: none"> • Indexeurs
Collecte	<ul style="list-style-type: none"> • Forwarders • Entrées modulaires • Réseau • HEC (Collecteur d'événements HTTP) • etc.
Gestion / Utilitaire	<ul style="list-style-type: none"> • CM • DS • LM • DMS • SHC-D

Aligner votre topologie sur les bonnes pratiques

Vous devrez garder votre cahier des charges et votre topologie à l'esprit pour choisir les principes de conception et bonnes pratiques qui conviennent à votre déploiement. Vous devez donc réfléchir aux bonnes pratiques seulement après avoir suivi les étapes 1 et 2 du processus de sélection des Architectures validées par Splunk ci-dessus.

Bonnes pratiques : recommandations propres à chaque couche

Vous trouverez ci-dessous les principes de conception et bonnes pratiques recommandés pour chaque couche de déploiement. Chaque principe de conception renforce une ou plusieurs bases fondamentales des AVS : disponibilité, performances, évolutivité, sécurité et gérabilité.

Recommandations pour la couche de recherche

PRINCIPES DE CONCEPTION / BONNES PRATIQUES (Vos spécifications déterminent les bonnes pratiques qui vous concernent)		BASES FONDAMENTALES D'AVS				
		DISPONIBILITÉ	PERFORMANCES	ÉVOLUTIVITÉ	SÉCURITÉ	GÉRABILITÉ
1	Maintenez la couche de recherche à proximité (en termes de réseau) de la couche d'indexation <i>Tout délai réseau intervenant entre la couche de recherche et la couche d'indexation aura un impact direct sur les performances de recherche.</i>		✓			
2	Évitez d'utiliser plusieurs search heads indépendantes <i>L'emploi de search heads indépendantes empêche le partage des artefacts Splunk créés par les utilisateurs. Elles sont aussi difficiles à faire évoluer sur le plan de l'utilisation des ressources de la couche de recherche. À moins qu'il ne faille spécifiquement créer des environnements de search heads indépendants, il existe de meilleures options, plus évolutives.</i>	✓		✓	✓	✓
3	Exploitez le clustering de search heads lorsque vous élargissez la couche de recherche <i>Un cluster de search heads permet de répliquer les</i>	✓		✓		

	<p>artefacts des utilisateurs dans l'ensemble du cluster et de planifier intelligemment la charge de recherche entre tous les membres du cluster. Il offre aussi une solution de haute disponibilité.</p>					
4	<p>Transmettez tous les logs internes des search heads à la couche d'indexation</p> <p>Toutes les données indexées doivent exclusivement être stockées dans la couche d'indexation. Nous évitons ainsi d'avoir à fournir un stockage hautes performances à la couche de search heads et cela simplifie la gestion. Remarque : cela s'applique également à tous les autres rôles Splunk.</p>		✓			✓
5	<p>Pensez à utiliser l'authentification LDAP autant que possible</p> <p>La gestion centralisée des identités des utilisateurs à des fins d'authentification est une bonne pratique globale en entreprise ; elle simplifie la gestion de votre déploiement Splunk et renforce sa sécurité.</p>				✓	✓
6	<p>Veillez à ce qu'il y ait suffisamment de cœurs pour satisfaire les besoins en recherches concurrentes</p> <p>L'exécution de chaque recherche mobilise un cœur de processeur. Sans cœur disponible pour exécuter une recherche, celle-ci sera mise en file d'attente, ce qui entraînera des retards pour l'utilisateur. Remarque : cela s'applique également à la couche d'indexation.</p>	✓	✓	✓		
7	<p>Utilisez autant que possible des fenêtres de recherche planifiées et homogénéisez la charge de recherches planifiées</p> <p>Les recherches planifiées sont souvent exécutées à des moments particuliers (à l'heure pile, 5/15/30 minutes</p>		✓	✓		

	<i>après l'heure, à minuit). En prévoyant la fenêtre dans laquelle votre recherche peut s'exécuter, vous évitez les pics de recherches simultanées.</i>					
9	<p>Limitez le nombre de clusters de search heads distincts pour éviter de surcharger la couche d'indexation</p> <p><i>La charge de recherche ne peut être régie automatiquement au sein d'un environnement de SH. Les SHC peuvent créer une charge de recherches simultanées supérieure aux capacités de la couche d'indexation (pairs de recherche). Ce point est aussi à prendre en compte pour bien planifier le nombre de search heads autonomes.</i></p>	✓		✓		
10	<p>Lors du montage des clusters de search heads, utilisez un nombre de nœuds impair (3, 5, 7, etc.).</p> <p><i>Le choix du Capitaine du SHC est assuré par un protocole reposant sur la majorité. Un nombre impair de nœuds évite qu'un SHC soit réparti en deux nombres de nœuds identiques en cas de panne réseau.</i></p>	✓				✓

Recommandations pour la couche d'indexation

PRINCIPES DE CONCEPTION / BONNES PRATIQUES (Vos spécifications déterminent les bonnes pratiques qui vous concernent)		BASES FONDAMENTALES				
		DISPONIBILITÉ	PERFORMANCES	ÉVOLUTIVITÉ	SÉCURITÉ	GÉRABILITÉ
1	<p>Activez des pipelines parallèles sur les serveurs qui le prennent en charge</p> <p><i>Les fonctions de parallélisation permettent l'exploitation de ressources système</i></p>		✓	✓		

	disponibles qui resteraient autrement inactives. Notez que les performances d'E/S doivent être adaptées avant d'activer les fonctions de parallélisation de l'ingestion.					
2	Pensez à utiliser des SSD pour les volumes contenant des données chaudes et les résumés Les SSD sont devenus abordables et éliminent toutes les limitations d'E/S qui sont souvent cause de mauvaises performances de recherche.		✓			
3	Maintenez la couche d'indexation à proximité (en termes de réseau) de la couche de recherche La latence réseau la plus faible possible aura un impact positif sur l'expérience des utilisateurs des recherches.		✓			
4	Utilisez la réplication d'index si vous devez garantir une HA des données historiques et des rapports La réplication d'index garantit la présence de plusieurs exemplaires de chaque événement dans le cluster, ce qui protège contre les défaillances des pairs de recherche. Ajustez le nombre de copies (facteur de réplication) en fonction de vos SLA.	✓				
5	Assurez une bonne hygiène d'importation des données (fins de ligne, extraction de l'horodatage, TZ, source, type de source et hôte sont correctement et explicitement définis pour chaque source de données) et établissez		✓	✓		✓

	<p>une surveillance continue à l'aide de la Console de surveillance</p> <p><i>Il s'avère que configurer explicitement les sources de données au lieu de s'appuyer sur les fonctions de détection automatique de Splunk offre d'importants avantages en termes de capacité d'ingestion des données et de latence d'indexation, en particulier dans les déploiements à grand volume.</i></p>					
6	<p>Pensez à configurer le paramètre de parallélisation des recherches par lots sur les indexeurs disposant d'un surplus de puissance de calcul</p> <p><i>L'exploitation des fonctions de parallélisation de la recherche peut avoir un impact significatif sur les performances de certains types de recherche et vous permet d'utiliser des ressources système qui resteraient autrement inactives.</i></p>		✓	✓		
7	<p>Veillez à ce que la répartition des données sur les nœuds d'indexeurs (pairs de recherche) soit équilibrée</p> <p><i>La répartition uniforme des événements et données sur les pairs de recherche est un facteur crucial pour les performances des recherches et la bonne application des politiques de conservation des données.</i></p>		✓	✓		✓
8	<p>Désactivez l'interface web sur les indexeurs dans les déploiements distribués/en cluster.</p> <p><i>Il n'est concrètement jamais utile d'accéder</i></p>		✓		✓	✓

	<i>directement à l'interface web des indexeurs.</i>					
9	<p>Pensez aux extensions technologiques prédéfinies de Splunk pour les sources de données courantes</p> <p><i>Plutôt que d'élaborer votre propre configuration pour assurer l'hygiène de l'importation des données provenant de sources connues, les extensions fournies par Splunk vous permettent d'obtenir plus rapidement des avantages et garantissent une implémentation optimale.</i></p>		✓			✓
10	<p>Surveillez les métriques critiques des indexeurs</p> <p><i>Splunk fournit une console de surveillance qui comprend des indicateurs clés portant sur les performances de votre couche d'indexation. Ils concernent notamment l'utilisation du processeur et de la mémoire. Ils comprennent aussi des métriques détaillées portant sur des composants internes de Splunk (processus, pipeline, files d'attente, recherches).</i></p>	✓	✓			

Recommandations pour la couche de collecte

PRINCIPES DE CONCEPTION / BONNES PRATIQUES (Vos spécifications déterminent les bonnes pratiques qui vous concernent)		BASES FONDAMENTALES				
		DISPONIBILITÉ	PERFORMANCES	ÉVOLUTIVITÉ	SÉCURITÉ	GÉRABILITÉ
1	Dans la mesure du possible, utilisez des UF pour transmettre les données. Le recours aux forwarders lourds doit être limité		✓			✓

	aux scénarios d'utilisation qui l'imposent. <i>AutoLB intégré, fonction de redémarrage autonome, configuration centralisée, faible utilisation de ressources</i>					
2	Utilisez au moins deux fois plus de pipelines de transmission intermédiaire que d'indexeurs lorsque vous canalisez de nombreux UF <i>Le multiplexage d'un grand nombre de forwarders de points de terminaison sur un petit nombre de forwarders intermédiaires nuit à la répartition uniforme des événements sur les indexeurs, ce qui affecte les performances de recherche. Ne déployez des forwarders intermédiaires qu'en cas d'absolue nécessité.</i>	✓	✓			
3	Pensez à sécuriser le trafic UF-IDX par SSL				✓	
4	Utilisez la fonction LB native de Splunk pour répartir les données dans la couche d'indexation <i>L'équilibrage des charges réseau n'est actuellement pas pris en charge <u>entre les forwarders et les indexeurs</u>.</i>	✓		✓		
5	Utilisez des serveurs syslog dédiés pour la collecte des données système <i>Les serveurs syslog peuvent assurer la persistance du trafic TCP/UDP sur un disque en fonction de la source, mais aussi appliquer une configuration de type de source permettant son traitement à l'aide d'un forwarder universel. Les redémarrages obligatoires des forwarders n'entraîneront aucune perte de données.</i>	✓				✓
6	Utilisez le HEC pour une collecte sans agent (plutôt que le protocole TCP/UDP natif) <i>Le Collecteur d'événements HTTP (HEC) est un service</i>	✓				✓

d'écoute qui permet de publier des événements via le protocole HTTP[S]. Il peut être activé directement sur les indexeurs ou configuré dans une couche de forwarder lourd, les deux desservis par un mécanisme d'équilibrage des charges.					
---	--	--	--	--	--

Recommandations pour la couche de gestion / utilitaire

PRINCIPES DE CONCEPTION / BONNES PRATIQUES (Vos spécifications déterminent les bonnes pratiques qui vous concernent)	BASES FONDAMENTALES				
	DISPONIBILITÉ	PERFORMANCES	ÉVOLUTIVITÉ	SÉCURITÉ	GÉRABILITÉ
1 Pensez à rassembler LM, CM, SHC-D et MC sur une même instance dans les environnements à petite échelle <i>Ces rôles de serveur utilisent très peu de ressources et sont des candidats idéaux à la colocation. Dans les clusters d'indexeurs plus volumineux, le CM peut avoir besoin d'un serveur dédié pour gérer efficacement le cluster.</i>					✓
2 Envisagez d'employer une instance distincte pour le DS dans les déploiements à moyenne à grande échelle <i>Une fois qu'un nombre significatif de forwarders est géré par le serveur de déploiement, les besoins en ressources augmentent jusqu'au point où un serveur dédié devient nécessaire pour maintenir le service.</i>					✓
3 Prévoyez plusieurs DS derrière un mécanisme de LB dans les très grands déploiements <i>Remarque : il peut s'avérer préférable de faire appel aux Services professionnels Splunk pour déployer et configurer correctement cette approche.</i>	✓		✓		✓

4	<p>Déterminez si le <code>phoneHomeIntervalInSecs</code> du DS peut être supérieur à la valeur par défaut de 60 secondes</p> <p><i>Un intervalle de contact plus long aura un effet positif sur l'évolutivité du DS.</i></p>			✓		
5	<p>Utilisez un DS dédié/sécurisé pour éviter l'exploitation des clients via le déploiement d'une application</p> <p><i>Toute personne ayant accès au Serveur de déploiement peut modifier la configuration de Splunk gérée par ce DS, et potentiellement déployer une application malveillante sur les points de terminaison des forwarders. Il est donc prudent de sécuriser correctement ce rôle.</i></p>				✓	
6	<p>Utilisez la Console de surveillance (MC) pour surveiller l'état de santé de votre déploiement et générer des alertes en cas de problème</p> <p><i>La console de surveillance fournit un ensemble prédéfini de solutions de surveillance propres à Splunk et inclut des alertes de plateforme extensibles qui peuvent vous avertir en cas de dégradation de votre environnement.</i></p>	✓	✓			✓

Récapitulatif et étapes suivantes

Ce livre blanc propose une présentation générale des Architectures validées par Splunk. Une architecture validée offre l'assurance que les exigences de votre entreprise seront satisfaites de la façon la plus économique, gérable et évolutive qui soit. Les AVS fournissent des bonnes pratiques et des principes de conception reposant sur les bases fondamentales suivantes :

- Disponibilité
- Performances
- Évolutivité
- Sécurité
- Gérabilité

Ce livre blanc aborde également le processus de sélection en trois étapes d'une Architecture validée par Splunk : 1) Définir les exigences, 2) Choisir une topologie et 3) Appliquer les principes de conception et bonnes pratiques. Maintenant que vous connaissez les nombreux avantages des Architectures validées par Splunk, nous espérons que vous vous sentez prêt à procéder au choix d'une topologie de déploiement adaptée à votre entreprise.

Étapes suivantes

Que faire une fois que vous avez choisi une architecture validée ? Voici les prochaines étapes de votre parcours vers un environnement fonctionnel :

Personnalisations

- Pensez à toutes les personnalisations dont la topologie que vous avez choisie aura éventuellement besoin pour satisfaire des exigences spécifiques.

Modèle de déploiement

- Choisissez votre modèle de déploiement : matériel local, virtuel, cloud.

Système

- Choisissez vos technologies (serveurs, stockage, systèmes d'exploitation) en fonction des exigences du système Splunk.

Dimensionnement

- Rassemblez toutes les données pertinentes dont vous aurez besoin pour dimensionner votre déploiement (ingestion des données, volume de recherche attendu, exigences de conservation des données, réplication, etc.). [Splunk Storage Sizing \(https://splunk-sizing.appspot.com/\)](https://splunk-sizing.appspot.com/) est un outil pratique pour réaliser cette estimation.

Personnel

- Évaluez vos besoins en personnel pour la mise en œuvre et la gestion de votre déploiement. Il s'agit d'un aspect essentiel de l'élaboration d'un Centre d'excellence Splunk.

Nous sommes là pour vous accompagner tout au long du processus de sélection d'une architecture validée et dans les étapes suivantes. N'hésitez pas à contacter votre équipe chargée de clientèle Splunk si vous avez la moindre question. Votre équipe chargée de clientèle aura accès à l'ensemble de la suite de ressources techniques et architecturales de Splunk et se fera un plaisir de vous donner des informations supplémentaires.

Bon Splunking !

Annexe


Cette section rassemble des informations de référence utilisées dans les AVS.






Annexe A : détails des bases fondamentales des AVS




Base fondamentale	Description	Objectifs principaux / Principes de conception
Disponibilité	Capacité d'un système à être continuellement opérationnel et capable de se rétablir en cas de panne ou de perturbation prévue ou imprévue.	<ol style="list-style-type: none"> 1. Éliminer les points de défaillance individuels / Ajouter de la redondance 2. Détecter les défaillances et pannes prévues et imprévues 3. Tolérer les pannes prévues et imprévues, idéalement de façon automatique 4. Planifier des mises à niveau roulantes
Performances	Capacité à utiliser efficacement les ressources disponibles pour maintenir un niveau de service optimal dans différents scénarios d'utilisation.	<ol style="list-style-type: none"> 1. Ajouter du matériel (calcul, stockage, mémoire) pour améliorer les performances. 2. Éliminer les goulots d'étranglement « à la base » 3. Exploiter tous les moyens de traitement simultané 4. Exploiter la localité (minimiser la dispersion des composants) 5. Optimiser pour le cas majoritaire (règle des 80/20) 6. Éviter les généralités inutiles 7. Décaler les calculs dans le temps (pré-calcul, calcul paresseux, calcul partagé/par lots) 8. Troquer la certitude et la précision contre du temps (randomisation, échantillonnage)
Évolutivité	Garantie que le système est pensé pour pouvoir s'agrandir dans toutes ses couches pour gérer efficacement des charges croissantes.	<ol style="list-style-type: none"> 1. S'agrandir verticalement et horizontalement 2. Séparer les composants fonctionnels qui doivent évoluer séparément 3. Minimiser les dépendances entre composants

Base fondamentale	Description	Objectifs principaux / Principes de conception
		<ol style="list-style-type: none"> 4. Concevoir en anticipant la croissance future aussi tôt que possible 5. Introduire une hiérarchie dans la conception globale du système
Sécurité	Garantie que le système est conçu pour protéger les données ainsi que les configurations et actifs tout en continuant à délivrer de la valeur.	<ol style="list-style-type: none"> 1. Concevoir un système sécurisé dès le départ 2. Employer des protocoles à la pointe de la technologie pour toutes les communications 3. Permettre un accès de haut niveau ou granulaire aux données d'événement 4. Employer une authentification centralisée 5. Mettre en œuvre des procédures d'audit 6. Réduire les surfaces d'attaque et d'utilisation malveillante
Gérabilité	Garantie que le système est conçu pour être contrôlable et gérable de manière centralisée pour toutes les couches.	<ol style="list-style-type: none"> 1. Fournir une fonction de gestion centralisée 2. Gérer le cycle de vie des objets de configuration (contrôle des sources) 3. Évaluer et surveiller/profiler l'utilisation des applications (Splunk) 4. Évaluer et surveiller l'état de santé du système

Annexe B : composants de topologie

Couche	Composant	Icône	Description	Remarques
Gestion	Serveur de déploiement (DS)		Le serveur de déploiement gère la configuration des forwarders.	Doit être déployé sur une instance dédiée. Peut être virtualisé pour une récupération simplifiée en cas de sinistre.

Couche	Composant	Icône	Description	Remarques
	Maître de licences (LM)		Le maître de licences est exigé par d'autres composants Splunk pour activer les fonctionnalités sous licence et suivre le volume quotidien d'ingestion de données.	Le rôle de maître de licences a de faibles besoins en capacité et en disponibilité, et il peut cohabiter avec d'autres fonctions de gestion. Peut être virtualisé pour une récupération simplifiée en cas de sinistre.
	Console de surveillance (MC)		La console de surveillance fournit des tableaux de bord permettant de surveiller l'utilisation et la santé de votre environnement. Elle comprend également différentes alertes de plateforme prédéfinies qui peuvent être personnalisées pour délivrer des notifications en cas de problème opérationnel.	Dans les environnements en cluster, la MC peut cohabiter avec le Nœud maître, en plus de la fonction de serveur de maître de licences et de déploiement dans les environnements sans cluster. Peut être virtualisé pour une récupération simplifiée en cas de sinistre.
	Maître de cluster (CM)		Le maître de cluster est le coordinateur indispensable pour toute activité dans un déploiement en cluster.	Dans les clusters présentant un grand nombre de buckets d'index (grand volume de données ou longue conservation), le maître de cluster nécessitera sans doute un serveur dédié. Peut être virtualisé pour une récupération facile en cas de désastre.
	Dépoyeur de cluster de search heads (SHC-D)		Le dépoyeur de cluster de search heads est nécessaire pour mettre en place un SHC et gérer la configuration Splunk déployée dans le cluster.	Le SHC-D n'est pas un composant runtime et utilise très peu de ressources système. Il peut cohabiter avec les autres rôles de gestion. <u>Remarque</u> : chaque SHC exige sa propre fonction de dépoyeur. Peut être virtualisé pour une récupération facile en cas de désastre.
Recherche	Search heads (SH)		La search head fournit une interface pour les utilisateurs de Splunk et coordonne l'activité de recherche planifiée.	Les search heads sont des instances Splunk dédiées dans les déploiements distribués. Les search heads peuvent être virtualisées pour faciliter la récupération en cas de sinistre, à condition qu'elles disposent des ressources

Couche	Composant	Icône	Description	Remarques
				processeur et mémoire adéquates.
	Cluster de search heads (SHC-D)		Un cluster de search heads est un pool d'au moins trois search heads en cluster. Il assure l'évolutivité horizontale de la couche de search heads et le basculement transparent des utilisateurs en cas de panne.	Les clusters de search heads nécessitent des serveurs dédiés présentant idéalement des caractéristiques système identiques. Les membres d'un cluster de search heads peuvent être virtualisés pour faciliter la récupération en cas de sinistre, à condition qu'ils disposent des ressources processeur et mémoire adéquates.
Indexation	Indexeur		Les indexeurs sont le cœur et l'âme de Splunk. Ils traitent et indexent les données entrantes et servent de pairs de recherche pour exécuter les requêtes initiées dans la couche de recherche.	Les indexeurs doivent toujours se trouver sur des serveurs dédiés dans les déploiements distribués ou en cluster. Dans un déploiement sur un seul serveur, l'indexeur fournira également l'interface de recherche et les fonctions de maître de licences. Les indexeurs fonctionnent de façon optimale sur des serveurs physiques locaux ou des machines virtuelles hautes performances dédiées, s'ils disposent des ressources adéquates.
Collecte de données	Forwarders et autres composants de collecte de données		Icône générique pour tout composant impliqué dans la collecte de données.	Cela inclut les forwarders universels et lourds, les entrées de données réseau et autres formes de collecte de données (HEC, Kafka, etc.)