splunk>

DOCUMENTO TÉCNICO

ARQUITECTURAS VALIDADAS DE SPLUNK

Índice de materias

Introducción

Las Arquitecturas Validadas de Splunk (SVA) son arquitecturas de referencia probadas para implementaciones estables, eficientes y repetibles de Splunk. Muchos de los clientes existentes de Splunk han disfrutado de una rápida adopción y expansión, que condujo a ciertos retos a medida que se intentaba la ampliación. Al mismo tiempo, los nuevos clientes de Splunk están buscando cada vez más directrices y arquitecturas certificadas para garantizar que sus implementaciones iniciales están construidas sobre cimientos sólidos. Las SVA se han desarrollado para ayudar a nuestros clientes con estas necesidades crecientes.

Independientemente de que sea un cliente nuevo o existente de Splunk, las SVA le ayudan a construir un entorno que es más fácil de mantener y en que la resolución de problemas es más sencilla. Las SVA están diseñadas para proporcionarle los mejores resultados posibles reduciendo al mismo tiempo su coste total de la propiedad. Del mismo modo, sus cimientos de Splunk tendrán una base arquitectónica repetible que le permitirá ampliar su implementación a medida que evolucionen sus necesidades con el tiempo.

Las SVA ofrecen opciones de topología que tienen en cuenta un amplio abanico de requisitos organizativos, de modo que puede comprender fácilmente y encontrar una topología que sea adecuada para sus requisitos. El proceso de selección de Arquitecturas Validadas de Splunk le ayudará a corresponder sus requisitos específicos con la topología que mejor se ajuste a las necesidades de su organización. Si es nuevo en Splunk, le recomendamos que implemente una Arquitectura Validada para su implementación inicial. Si ya es uno de nuestros clientes, le recomendamos que explore la opción de alinearse con una topología de Arquitectura Validada. A no ser que tenga requisitos exclusivos que hagan necesaria la construcción de una arquitectura personalizada, es muy probable que una Arquitectura Validada cumpla sus requisitos manteniendo la rentabilidad.

Este documento técnico le proporciona una visión general de las SVA. Dentro de este documento técnico encontrará los recursos que necesita para pasar por el proceso de selección de SVA, incluyendo el cuestionario de requisitos, diagramas de topología de implementación, principios de diseño y directrices generales.

Si necesita ayuda para implementar una Arquitectura Validada de Splunk, póngase en contacto con los Servicios Profesionales de Splunk (https://www.splunk.com/en_us/support-and-services/splunk-services.html).

Estructura del documento

Las SVA se desglosan en tres áreas de contenido principales:

1. Topologías de indexado y búsqueda
2. Componentes de la arquitectura de recopilación de datos
3. Principios de diseño y prácticas recomendadas

La indexación y la búsqueda cubren los niveles arquitectónicos que proporcionan las funciones indexado y búsqueda principales del núcleo de una implementación de Splunk. La sección del componente de recopilación de datos le guía por la selección del mecanismo de recopilación de datos adecuado para sus requisitos.

Los principios de diseño y las prácticas recomendadas se aplican a su arquitectura como un todo y le ayudarán a elegir correctamente cuando prepare los detalles de su implementación.

Razones para utilizar arquitecturas validadas de Splunk

La implementación de una arquitectura validada le dará la capacidad de diseñar e implementar Splunk con mayor confianza. Las SVA le ayudarán a resolver algunos de los retos más habituales que afrontan las organizaciones, incluyendo:

Rendimiento

- Las organizaciones quieren ver mejoras en el rendimiento y la estabilidad.

Complejidad

- Las organizaciones a veces caen en las trampas de las implementaciones construidas a medida, especialmente cuando han crecido con demasiada rapidez u orgánicamente. En dichos casos, se habría podido introducir la complejidad innecesaria. Esta complejidad podría convertirse en un serio escollo al intentar la ampliación.

Eficiencia

- Para derivar el máximo de los beneficios de la implementación de Splunk, las organizaciones deben mejorar la eficiencia de las operaciones y acelerar el tiempo de generación de valor.

Coste

- Las organizaciones buscan formas de reducir los costes totales de la propiedad (TCO) y cumplir al mismo tiempo el resto de sus requisitos.

Agilidad

- Las organizaciones se tendrán que adaptar a los cambios a medida que se amplíen y crezcan.

Mantenimiento

- La optimización del entorno es a menudo necesaria para reducir los esfuerzos de mantenimiento

Capacidad de ampliación

- Las organizaciones deben tener la capacidad de ampliarse de forma eficiente y transparente.

Verificación

- Las partes interesadas de la organización quieren la garantía de que su implementación de Splunk está construida sobre prácticas recomendadas.

Pilares de las Arquitecturas Validadas de Splunk

Las Arquitecturas Validadas de Splunk están construidas sobre los siguientes pilares fundamentales. Para obtener más información sobre estos pilares de diseño, consulte el Apéndice A a continuación.

DISPONIBILIDAD	RENDIMIENTO	CAPACIDAD DE AMPLIACIÓN	SEGURIDAD	CAPACIDAD DE GESTIÓN
El sistema tiene un **funcionamiento continuo** y es capaz de recuperarse de cortes o interrupciones planificados y no planificados.	El sistema puede **mantener un nivel óptimo de servicio** bajo diversos patrones de uso.	El sistema está diseñado para ampliarse en todos los niveles, lo que le permite **tratar las cargas de trabajo aumentadas de manera efectiva**.	El sistema está **diseñado para proteger los datos, las configuraciones y los activos** con una entrega de valor continua.	El sistema es **operativo y gestionable de forma centralizada en todos sus niveles**.

Estos pilares sustentan directamente el Servicio **Gestión y cobertura de plataforma** en el modelo Splunk Center Of Excellence (Centro de distinción de Splunk).

Qué esperar de las Arquitecturas validadas de Splunk

Tenga en cuenta que las SVA no incluyen las tecnologías de implementación ni el cálculo del tamaño de la implementación El razonamiento de este punto es el siguiente:

- Las tecnologías de implementación, como los sistemas operativos y el hardware de servidores, se consideran opciones de implementación en el contexto de las SVA. Los diversos clientes tendrán elecciones diferentes, de modo que la generalización no es posible fácilmente.
- Los tamaños de las implementaciones requieren una evaluación del volumen de introducción de datos, los tipos de los datos, los volúmenes de las búsquedas y los casos de uso de las búsquedas, que tienen tendencia a ser muy específicos de los clientes y normalmente no tienen influencia en la arquitectura de implementación fundamental en sí. Las herramientas de cálculo de tamaño pueden ayudar con este proceso una vez haya establecido su arquitectura de implementación. Cálculo de tamaño de almacenamiento de Splunk (https://splunk-sizing.appspot.com/) es una de las herramientas disponibles.

Las SVA proporcionan:	**Las SVA no proporcionan:**
✅ Opciones de implementación agrupadas en clústeres o no. ✅ Diagramas de la arquitectura de referencia. ✅ Directrices para ayudarle a seleccionar la arquitectura adecuada para usted. ✅ Recomendaciones específicas de nivel. ✅ Prácticas recomendadas para construir su implementación de Splunk.	❌ Elecciones de implementación (sistema operativo, desde cero frente a virtual frente a nube, etc.). ❌ Cálculo del tamaño de la implementación. ❌ Una aprobación preceptiva de su arquitectura. Nota: Las SVA proporcionan recomendaciones y directrices, de modo que puede tomar la decisión para su organización en última instancia. ❌ Una sugerencia de topología para cada posible escenario de implementación. En algunos casos, los factores exclusivos pueden requerir que se implemente una arquitectura personalizada. Los expertos de Splunk están disponibles para ayudarle con cualquier solución personalizada que necesite. Si ya es cliente, póngase en contacto con su Equipo de cuenta de Splunk. Si es nuevo en Splunk, puede ponerse en contacto con nosotros aquí (https://www.splunk.com/en_us/talk-to-sales.html).

Funciones y responsabilidades

Las Arquitecturas Validadas de Splunk son altamente relevantes para las preocupaciones de los encargados de la toma de decisiones y los administradores. Los arquitectos empresariales, los asesores, los administradores de Splunk y los proveedores de servicios gestionados deberían implicarse en el proceso de selección de las SVA. Encontrará una descripción de cada una de estas funciones a continuación:

Función	**Descripción**
Arquitectos empresariales	Responsables de que la arquitectura de las implementaciones de Splunk cumplan las necesidades empresariales.

Función	Descripción
Asesores	Responsables de proporcionar servicios para la arquitectura, el diseño y la implementación de Splunk.
Ingenieros de Splunk	Responsables de gestionar el ciclo de vida de Splunk.
Proveedores de servicios administrados	Entidades que implementan y ejecutan Splunk como un servicio para clientes.

Descripción general del proceso de selección de Arquitecturas Validadas de Splunk

El proceso de selección de Arquitecturas Validadas de Splunk le ayudará a identificar la arquitectura más sencilla y optimizada que cumpla todas las necesidades de su organización.

Pasos del proceso de selección	Objetivos	Consideraciones
Paso 1: Definir los requisitos para: a) Indexación y búsqueda b) Mecanismo(s) de recopilación de datos	*Defina los requisitos.*	• Los responsables de la toma de decisiones, las partes interesadas y los administradores deberían colaborar en la identificación y definición de los requisitos de su organización. • Si ya tiene una implementación vigente, puede evaluar su arquitectura actual para ver si debería trasladarse a un modelo validado. *Para ver un cuestionario que le ayudará a definir sus requisitos, consulte el Paso 1 que aparece a continuación.*
Paso 2: Seleccionar una topología para: a) Indexación y búsqueda b) cada mecanismo de recopilación de datos	*Seleccione una topología que coincida con los requisitos identificados.*	• Seleccione una topología que coincida con sus requisitos. • Mantenga las cosas sencillas y de acuerdo con el SVA, de modo que pueda ver el camino más sencillo a la ampliación. *Para ver diagramas y descripciones de opciones de topología, consulte el Paso 2 a continuación.*

Paso 3: Aplicar principios de diseño y prácticas recomendadas	*Priorice sus principios de diseño y revise prácticas de recomendadas de implementación específicas del nivel.*	• Cada principio de diseño afianza uno o más pilares de las Arquitecturas Validadas de Splunk. • Priorizará los principios de diseño según las necesidades de su organización. • Las recomendaciones específicas del nivel guiarán su topología de implementación. *Para ver un desglose de los principios de diseño, consulte el Paso 3 a continuación.*

Paso 1a: Definir sus requisitos de indexación y búsqueda

Para seleccionar la topología de implementación apropiada, tendrá que profundizar en sus requisitos. Una vez haya definido sus requisitos, podrá seleccionar la forma más sencilla y rentable de implementar Splunk. A continuación encontrará un cuestionario que le ayudará a definir áreas de requisitos clave para los niveles de indexación y búsqueda de su implementación.

El cuestionario de requisitos se centra en áreas que tendrán una repercusión directa en su topología de implementación. Por lo tanto, recomendamos encarecidamente que guarde sus respuestas a las siguientes preguntas antes de seleccionar una topología en el siguiente paso.

Cosas a tener en cuenta

Revise sus casos de uso

A medida que defina sus requisitos, debería pensar el los casos de uso a los que va a ir destinada su infraestructura de Splunk. Por ejemplo, la topología de un caso de uso de operaciones de desarrollo es a menudo más sencilla que un caso de uso crítico de la misión (aunque no en todos los casos). Debería considerar completamente casos de uso que conlleven:

- Búsqueda
- Disponibilidad
- Requisitos de cumplimiento (esto es especialmente importante si necesita tener un 100% de fidelidad y disponibilidad de los datos en todo momento)
- Otros escenarios de casos de uso específicos de su organización

Dependiendo de sus escenarios de casos de uso, su implementación podría tener que proporcionar características arquitectónicas adicionales.

Piense en el crecimiento futuro

Tendrá que pensar en sus necesidades inmediatas para definir sus requisitos. No obstante, también debería tener en cuenta el crecimiento y la capacidad de ampliación futuros. La ampliación de su implementación podría requerir desembolsos, personal adicional u otros recursos que sería bueno que empezase a planificar hoy mismo.

Categorías de topologías

Lo siguiente es clave en las categorías de topologías de la SVA. Estas categorías se utilizan en el cuestionario siguiente. También encontrará referencias a estas categorías en los siguientes pasos del proceso de selección de la SVA.

Categorías de nivel de indexado

Código de categoría	Explicación
S	La categoría "S" indica el indexador de una implementación de Splunk con un único servidor
D	La categoría "D" indica la necesidad de un nivel de indexadores distribuidos con al menos 2 indexadores
C	La categoría "C" indica la necesidad de un nivel de indexadores agrupados en clústeres (se requiere replicación de datos)
M	La categoría "M" indica la necesidad de un nivel de indexadores agrupados en clústeres con sitios múltiples

Categorías de nivel de búsqueda

Código de categoría	Explicación
1	La categoría "1" indica que una única cabeza de búsqueda podría cumplir los requisitos
2	La categoría "2" indica que se requieren varias cabezas de búsqueda para cumplir los requisitos
3	La categoría "3" indica que se requiere una agrupación en clúster de cabezas de búsqueda para cumplir los requisitos
4	La categoría "4" indica que se requiere una agrupación en clúster de cabezas de búsqueda (SHC) que abarca varios sitios (una SHC "extendida") para cumplir los requisitos
+10	La categoría "+10" indica que se requiere una cabeza de búsqueda (agrupación en clúster) dedicada para dar cobertura para la aplicación de seguridad empresarial. Agregue 10 a la categoría de la topología del nivel de búsqueda y lea detenidamente en la descripción de la topología los requisitos específicos para esta aplicación.

Cuestionario 1: Definición de sus requisitos de para los niveles de indexación y búsqueda

♦ *Consulte la clave anterior para una explicación de los códigos de las categorías de las topologías. Si responde "sí" a varias preguntas, utilice el código de la categoría de la topología para la pregunta con el número más alto.*

Nº	Pregunta	Consideraciones	Repercusión sobre la topología	Categoría de topología de nivel de indexador ♦	Categoría de topología de nivel de búsqueda ♦
1	¿Es su introducción de datos prevista inferior a ~300 GB/día?	Considere un crecimiento a corto plazo en la introducción diaria (~6-12 meses)	Candidato para una implementación de un único servidor, dependiendo de las preguntas relacionadas	S	1

Nº	Pregunta	Consideraciones	Repercusión sobre la topología	Categoría de topología de nivel de indexador ♦	Categoría de topología de nivel de búsqueda ♦
			con la disponibilidad		
2	¿Requiere una alta disponibilidad de la recopilación/ indexado de los datos?	Si no tiene intención de utilizar Splunk para la supervisión de casos de uso que requieran una introducción de datos continua, una interrupción temporal del flujo de datos entrantes podría ser aceptable, siempre que no se pierdan datos de registro.	Requiere una implementación distribuida para dar cobertura a la introducción continua	D	1
3	Suponiendo que una cabeza de búsqueda realice una búsqueda: ¿Tienen sus datos que poder buscarse completamente en todo momento (por ej. no puede permitirse ningún impacto en la integridad de los resultados de las búsquedas?	Si su caso de uso está calculando mediciones de rendimiento y supervisión de uso general empleando funciones agregadas, por ejemplo, una interrupción aislada del indexador podría no afectar materialmente el cálculo de datos estadísticos sobre un número elevado de incidencias. Si su caso de uso es la auditoría de seguridad y la detección de amenazas, los puntos ciegos en los resultados de las búsquedas son muy probablemente poco deseables.	Requiere indexadores agrupados en clústeres con un factor de replicación de al menos dos (2). Nota: aunque un factor de replicación de 2 proporciona una protección mínima contra el fallo de nodo de indexador único, el factor de replicación recomendado (y predeterminado) es de 3.	C	1
4	¿Tiene centros de datos	Los requisitos de recuperación de	El funcionamiento	M	2

Nº	Pregunta	Consideraciones	Repercusión sobre la topología	Categoría de topología de nivel de indexador ♦	Categoría de topología de nivel de búsqueda ♦
	múltiples y requiere la recuperación automática de su entorno de Splunk en caso de una interrupción del centro de datos?	desastres pueden dictar el funcionamiento continuo de dos instalaciones (activas/activas) o prescribir objetivos RTO/RPO para la recuperación de desastres manual	continuo requerirá la agrupación en clústeres de los indexadores en varios emplazamientos y al menos dos cabezas de búsqueda activas para garantizar la protección contra los fallos tanto en el nivel de introducción/ indexación de los datos como en el nivel de las búsqueda.		
5	Suponiendo una introducción de datos continua y sin pérdidas, ¿requiere alta disponibilidad para el nivel de búsqueda de cara al usuario?	Si se está utilizando Splunk para la supervisión continua casi en tiempo real, las interrupciones en el nivel de búsqueda no son tolerables probablemente. Esto puede ser cierto o no para otros casos de uso.	Requiere cabezas de búsqueda redundantes, y potencialmente la agrupación en clústeres de las cabezas de búsqueda	D/C/M	3
6	¿Necesita dar cobertura a un gran número de usuarios simultáneos y/o una carga de trabajo de búsqueda significativamente programada?	Los requisitos para más de ~50 usuarios/ búsquedas simultáneos normalmente requieren la ampliación horizontal del nivel de búsqueda	Puede ser necesaria una topología que utilice una agrupación en clúster de cabezas de búsqueda en el nivel de búsqueda	D/C/M	3
7	En un entorno de múltiples centros de datos, ¿necesita que se sincronicen los artefactos de los	Esto decidirá si los usuarios disfrutan de una experiencia vigente y coherente en el caso de una	Requiere una agrupación en clúster "extendida" de las cabezas de búsqueda entre sitios con una	M	4

Nº	Pregunta	Consideraciones	Repercusión sobre la topología	Categoría de topología de nivel de indexador ♦	Categoría de topología de nivel de búsqueda ♦
	usuarios (búsquedas, paneles y otros objetos de conocimiento) entre sitios?	interrupción del sitio.	configuración apropiada. **Importante:** Aunque una SHC extendida puede mejorar la disponibilidad para los usuarios durante un fallo de sitio completo, no puede garantizarse que todos los artefactos se repliquen entre ambos sitios en todo momento. Esto puede afectar aplicaciones específicas que dependen de artefactos coherentes y vigentes, como la Aplicación Splunk para la seguridad empresarial. La agrupación en clústeres de cabezas de búsqueda por sí sola no puede proporcionar una solución DR completa. Otros beneficios para SHC sí se aplican.		
8	¿Tiene intención de implementar la Aplicación Splunk para la seguridad empresarial (ES)?	Asegúrese de leer y comprender las limitaciones específicas a las que está sujeta la Aplicación Splunk para la seguridad empresarial según se documenta con cada topología.	ES requiere un entorno de cabezas de búsqueda exclusivo (ya sea autónomo o agrupado en clúster).	D/C/M	+10

Nº	Pregunta	Consideraciones	Repercusión sobre la topología	Categoría de topología de nivel de indexador ♦	Categoría de topología de nivel de búsqueda ♦
9	¿Tiene un entorno distribuido geográfica-mente que esté sujeto a normativas de custodia de datos?	Las normativas de algunos países no permiten que los datos generados dentro del país abandonen los sistemas de ese país.	Dichas normativas prohíben la implementación de un nivel de indexado central de Splunk y requieren que se desarrolle una arquitectura personalizada por parte de una colaboración entre Splunk/socio y el cliente que tenga en cuenta los detalles de dicha implementación en profundidad. En otras palabras, no hay una SVA para cumplir este requisito.	Personalizada	Personalizada
10	¿Tiene directrices de seguridad altamente restrictivas que impiden la ubicación conjunta de fuentes de datos de registro específicas en servidores/index adores compartidos?	Es posible que no se permita que los datos de registro altament e confidenciales se ubiquen conjuntamente con conjuntos de datos de riesgo inferior en el mismo sistema físico o dentro de la misma zona de red en base a directrices corporativas.	Se necesitan entornos de indexado independientes y múltiples, potencialmente con un nivel de búsqueda híbrido compartido. Esto va más allá del ámbito de las SVA y requiere un desarrollo arquitectónico personalizado.	Personalizada	Personalizada

Cómo determinar el código de la categoría de su topología

En base a sus respuestas al cuestionario de requisitos anterior, llegará a un indicador de categoría de topología combinado que le permitirá identificar la mejor topología para sus necesidades. Se proporcionan instrucciones y ejemplos a continuación.

Instrucciones

1. Escriba las preguntas a las que respondió afirmativamente.

2. Si respondió "sí" a varias preguntas, siga la recomendación de topología para la pregunta con el número más alto. Si ve varias opciones de topología (por ejemplo, "D/C/M"), mire las preguntas anteriores para determinar qué opción es la que mejor se ajusta a su caso.
3. Su código de categoría de topología comienza por la letra que representa el nivel del indexador (por ejemplo, "C" o "M"). Esta letra estará seguida del número que representa el nivel de búsqueda (por ejemplo, "1" o "13").

Ejemplo Nº1

Digamos que respondió afirmativamente a las preguntas 3, 5 y 8. Llegará a una categoría de topología de "C13", lo que indica la necesidad de un nivel de indexado agrupado en clústeres con dos clústeres de cabezas de búsqueda.

Ejemplo Nº2

Ahora digamos que respondió afirmativamente solo a la pregunta 1. Llegará a una categoría de topología de "S1", lo que indica una implementación de un único servidor de Splunk como su topología ideal.

Paso 2a: Elegir una topología para indexación y búsqueda

Las topologías se dividen generalmente en implementaciones no agrupadas y agrupadas en clústeres. Las implementaciones no agrupadas en clústeres requieren la mejor cantidad de componentes diferenciados y tienen excelentes características de ampliación. Tenga en cuenta que aunque las implementaciones no agrupadas en clústeres vienen con una disponibilidad y funciones de recuperación de desastres reducidas, esta opción de implementación podría aún ser una buena elección para su organización.

Recuerden: El objetivo principal del proceso de selección de la SVA es permitirle construir lo que necesita sin aportar componentes innecesarios.

Nota

Aunque podría elegir implementar una topología que proporcione beneficios adicionales más allá de sus necesidades inmediatas, tenga en cuenta que esto desembocará probablemente en costes innecesarios. Y lo que es más, la incorporación de complejidad adicional es a menudo contraproducente en la eficiencia operativa.

Nota importante sobre los diagramas de topologías

Los iconos de los diagramas de topologías representan **funciones operativas de Splunk** y no implican una infraestructura dedicada para ejecutarlas. Consulte el Apéndice para obtener directrices sobre qué funciones de Splunk pueden ser colocadas en la misma infraestructura o servidor.

Uso de su código de categoría de topología

Antes de seleccionar una opción de topología, se recomienda encarecidamente que complete el cuestionario de requisitos para determinar su código de categoría de topología. Si aún no ha hecho esto, vuelva y complete el paso anterior. Una vez tenga su código de categoría de topología podrá identificar la opción de implementación que mejor se ajuste a sus requisitos declarados.

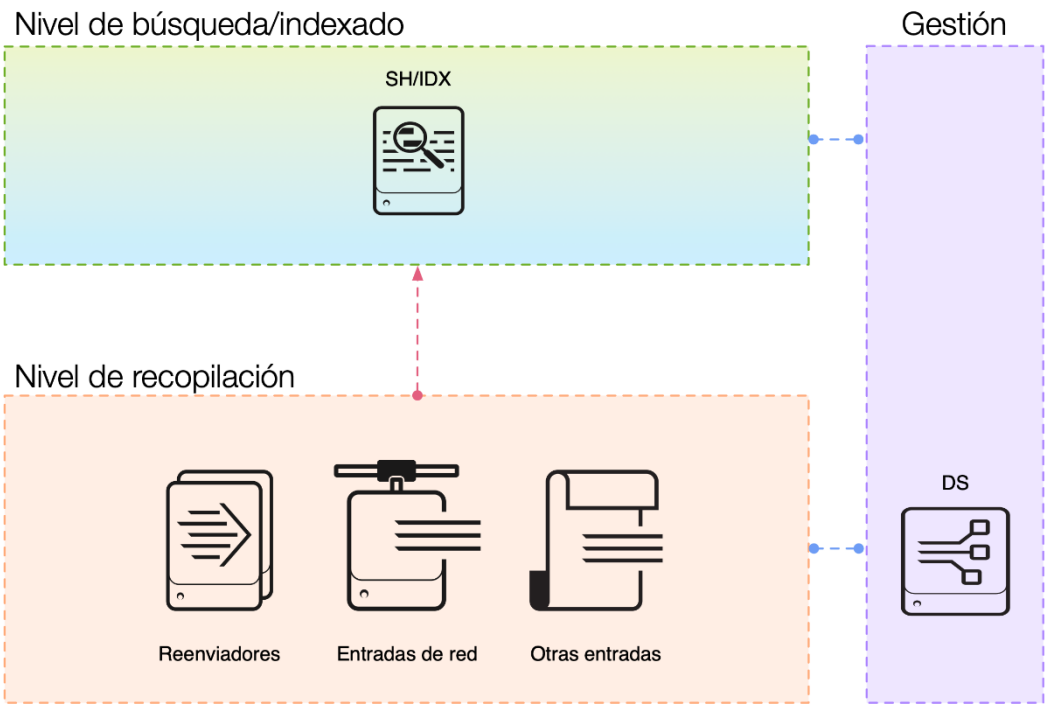
Opciones de implementación no agrupadas en clústeres

A continuación encontrará las siguientes opciones de topología:

Tipo de implementación	Código(s) de categoría de topología
Implementación de un solo servidor	S1
Implementación no agrupada en clúster distribuida	D1 / D11

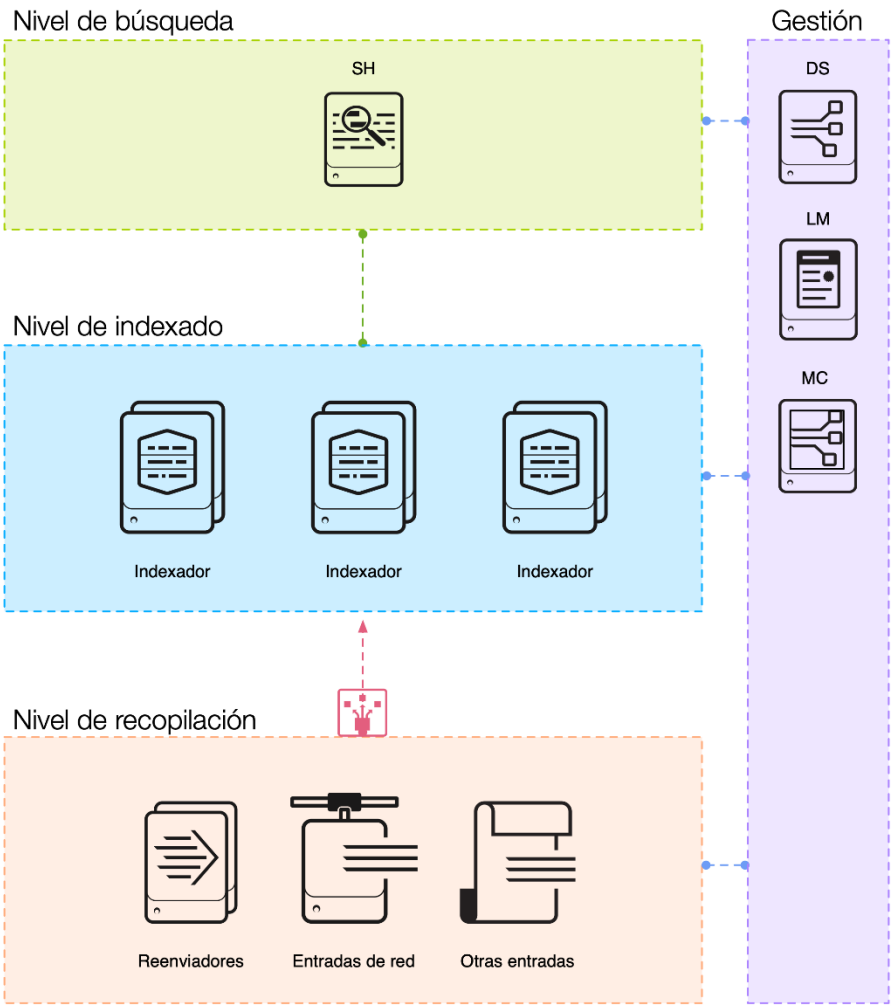
Para consultar una explicación de los componentes de topología, consulte el Apéndice "B" a continuación.

Implementación de un solo servidor (S1)



Descripción de la implementación de un solo servidor (S1)	Limitaciones
Esta topología de implementación le proporciona una solución muy rentable si su entorno cumple todos los siguientes criterios: a) no tiene requisitos de proporcionar alta disponibilidad o recuperación de desastres automática para su implementación de Splunk; b) su introducción de datos diaria está por debajo de los 300 GB/día; c) tiene un reducido número de usuarios con casos de uso de búsqueda no críticos. Esta topología se utiliza normalmente para casos de uso más pequeños que no son críticos para la actividad comercial (a menudo departamentales en su naturaleza). Los casos de uso apropiados incluyen entornos de prueba de incorporación de datos, casos de uso de operaciones de desarrollo pequeñas, la prueba de aplicaciones y entornos de integración y escenarios similares. Los beneficios principales de esta topología incluyen una capacidad de gestión sencilla, un buen rendimiento de búsqueda para volúmenes de datos pequeños y un coste total de propiedad fijo.	• Sin alta disponibilidad para la búsqueda/indexado • La capacidad de ampliación está limitada por la capacidad del hardware (ruta de migración directa a una implementación distribuida)

Implementación no agrupada en clúster distribuida (D1 / D11)



Descripción de la implementación no agrupada en clúster distribuida (D1 / D11)	Limitaciones
Tendrá que pasarse a una topología distribuida en cualquiera de las situaciones siguientes: a) el volumen de datos diario que tiene que enviar a Splunk supera la capacidad de una implementación de un único servidor; b) desea o necesita proporcionar una introducción de datos de alta disponibilidad. La implementación de múltiples indexadores independientes le permitirá ampliar su capacidad de indexación linealmente y aumentar implícitamente la disponibilidad para la introducción de datos. El coste total de propiedad aumentará de forma predecible y lineal a medida que incorpore nodos de indexadores. La introducción recomendada del componente Consola de supervisión (MC) le permite supervisar el estado y la capacidad de su implementación distribuida. Adicionalmente, la MC proporciona un sistema de alerta centralizado, de modo que se le notificarán las condiciones de mal estado de su implementación. Las cabezas de búsqueda tendrán que configurarse manualmente con la lista de puntos de búsqueda disponibles cada vez que se incorporen nuevos indexadores. **Nota para clientes de ES:** Si su código de categoría es D1 (por ej. tiene intención de implementar la Aplicación Splunk para la seguridad empresarial), se requiere una cabeza de búsqueda **única en exclusiva** para implementar la aplicación (esto no se ilustra en el diagrama de topologías). El nivel de recopilación tiene que configurarse con la lista de indexadores de destino (a través de un servidor de implementación) cada vez que se incorporen nuevos indexadores. Esta topología de implementación puede ampliarse linealmente a más de 1000 nodos de indexadores y por ello puede dar cobertura a una introducción de datos y volúmenes de búsquedas extremadamente altos. Se puede mantener el rendimiento de las búsquedas entre conjuntos de datos de gran tamaño a través de una ejecución paralela de búsquedas entre muchos indexadores (asignar/reducir). Aunque no se desglosa como una topología separada, un clúster de cabezas de búsqueda puede utilizarse para aumentar la capacidad de las búsquedas en el nivel de búsqueda (consulte el nivel de búsqueda en la topología C3/C13).	• Sin alta disponibilidad para el nivel de búsqueda • Alta disponibilidad limitada para el nivel de indexado, el fallo de nodos podría provocar resultados de búsquedas incompletos para búsquedas históricas

Opciones de implementación agrupadas en clúster

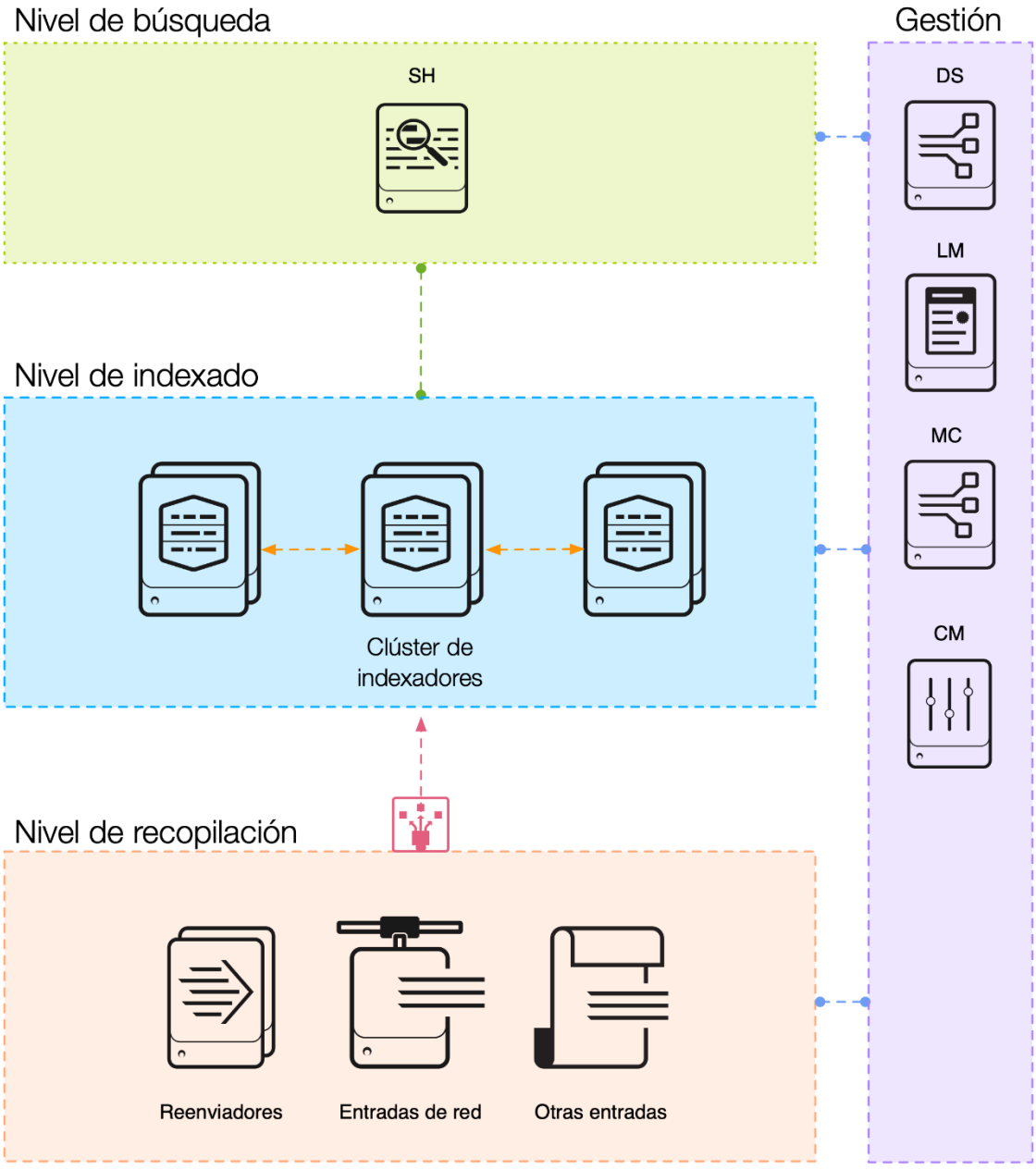
A continuación encontrará las siguientes opciones de topología:

Tipo de implementación	Código(s) de categoría de topología
Implementación agrupada en clúster distribuida - Único sitio	C1 / C11

Tipo de implementación	Código(s) de categoría de topología
Implementación agrupada en clúster distribuida + SHC - Único sitio	C3 / C13
Implementación agrupada en clúster distribuida - Varios sitios	M2 / M12
Implementación agrupada en clúster distribuida + SHC - Varios sitios	M3 / M13
Implementación agrupada en clúster distribuida + SHC - Varios sitios	M4 / M14

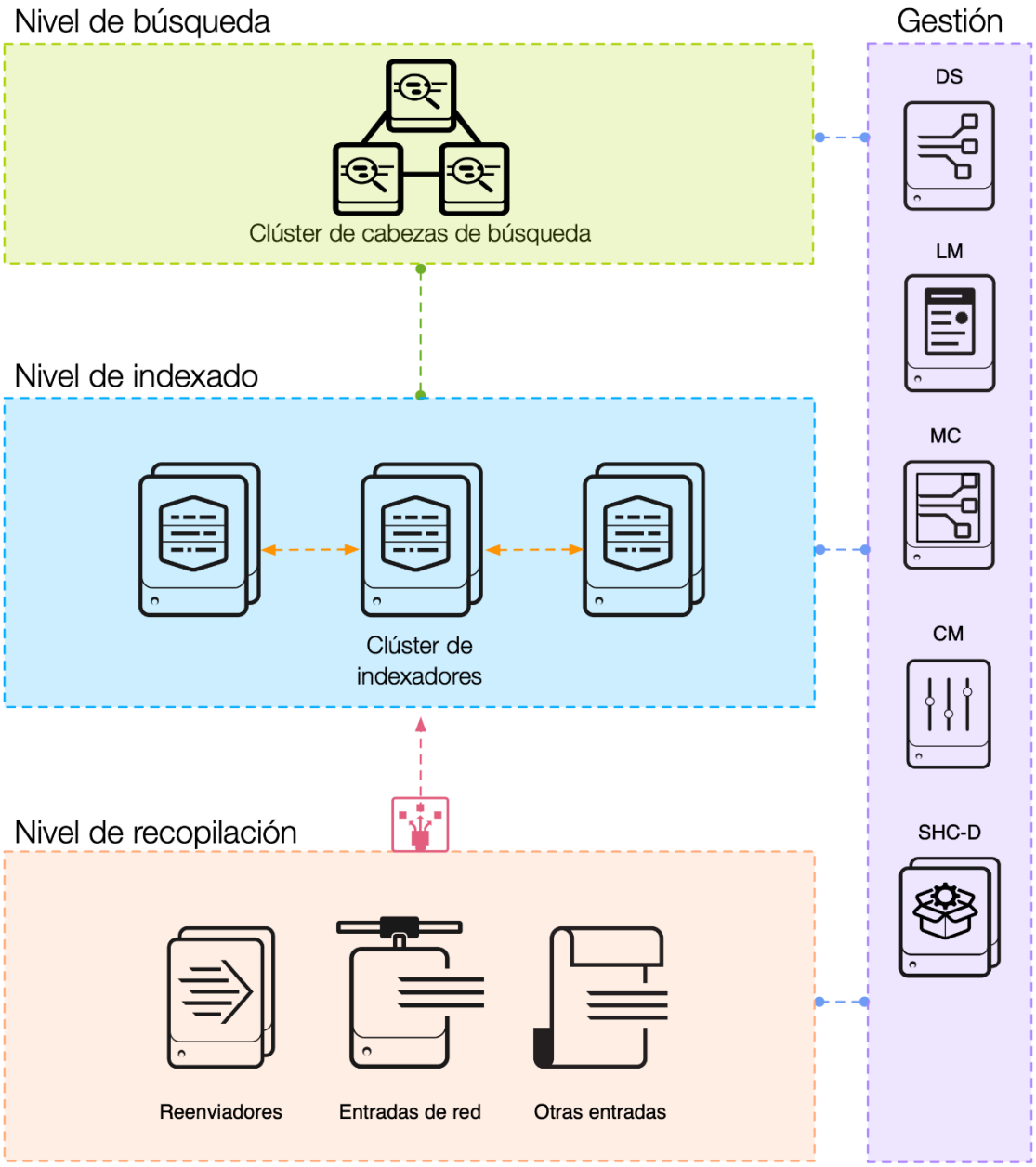
Para consultar una explicación de los componentes de topología, consulte el Apéndice "B" a continuación.

Implementación agrupada en clúster distribuida - Único sitio (C1 / C11)



Descripción de la implementación agrupada en clúster distribuida - Único sitio (C1 / C11)	Limitaciones
Esta topología presenta la agrupación en clústeres de indexadores en conjunción con una directriz de replicación de datos configurada apropiadamente. Esto proporciona una alta disponibilidad de los datos en caso de fallo del nodo del punto del indexador. No obstante, debe ser consciente de que esto se aplica únicamente al nivel de indexado y no protege frente a fallos de cabezas de búsqueda. **Nota para clientes de ES:** Si su código de categoría es C11 (por ej. tiene intención de implementar la Aplicación Splunk para la seguridad empresarial), se requiere una cabeza de búsqueda **única en exclusiva** para implementar la aplicación (esto no se ilustra en el diagrama de topologías). Esta topología requiere un componente de Splunk adicional denominado Maestro de clúster (CM). El CM es responsable de la coordinación y la aplicación de la directriz de replicación de datos configurada. El CM también sirve como fuente de autorización para los puntos de clúster disponibles (indexadores). La configuración de las cabezas de búsqueda se simplifica configurando el CM en vez de los puntos de búsqueda individuales. Tiene la opción de configurar el nivel de reenvío para detectar los indexadores disponibles a través del CM. Esto simplifica la gestión del nivel de reenvío. Tenga en cuenta que los datos se replican en el clúster de manera no determinista. No tendrá el control sobre la ubicación donde se almacenan las copias solicitadas de cada evento. De forma adicional, aunque la capacidad de ampliación es lineal, existen limitaciones en referencia al tamaño total del clúster (~50 PB de datos con capacidad de búsqueda bajo condiciones ideales). Recomendamos la implementación de la Consola de supervisión (MC) para supervisar el estado de su entorno de Splunk.	• Sin alta disponibilidad para el nivel de búsqueda • El número total de depósitos exclusivos en el clúster del indexador está limitado a 5MM (V6.6+), 15MM depósitos en total • Si capacidad de recuperación de desastres automática en caso de interrupción del centro de datos

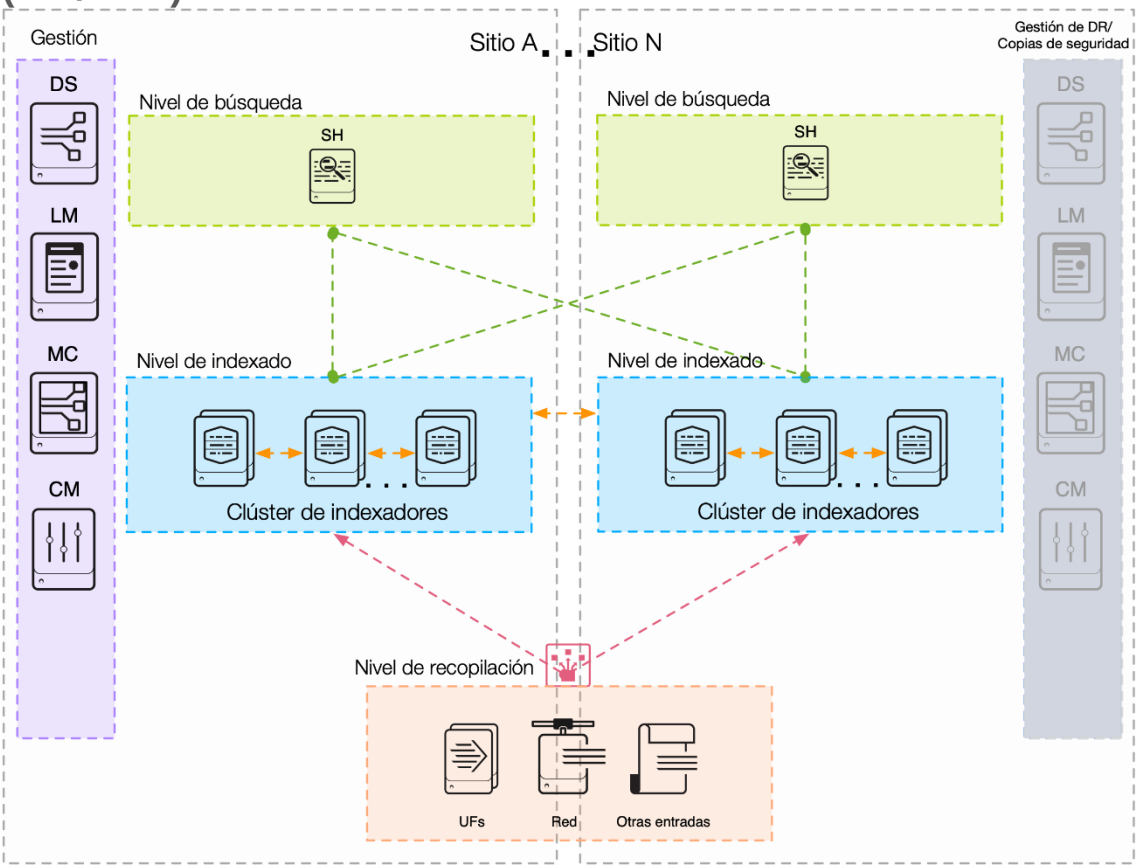
Implementación agrupada en clúster distribuida + SHC - Único sitio (C3 / C13)



Descripción de la implementación agrupada en clúster distribuida + SHC - Único sitio (C3 / C13)	Limitaciones
Esta topología añade capacidad de ampliación horizontal y elimina el punto de fallo único del nivel de búsqueda. Se requiere un mínimo de tres cabezas de búsqueda para implementar una SHC Para gestionar la configuración de la SHC, se requiere un componente de Splunk adicional denominado Implementador de clúster de cabezas de búsqueda para cada SHC. Este componente es necesario para implementar cambios en los archivos de configuración del clúster. El Implementador de clúster de cabezas de búsqueda no tiene requisitos de alta disponibilidad (sin función en tiempo de ejecución).	• Si capacidad de recuperación de desastres en caso de interrupción del centro de datos • ES requiere una SH/SHC exclusiva • Se admite la gestión de una implementación de ES en SHC, pero supone un reto (conlleva PS)

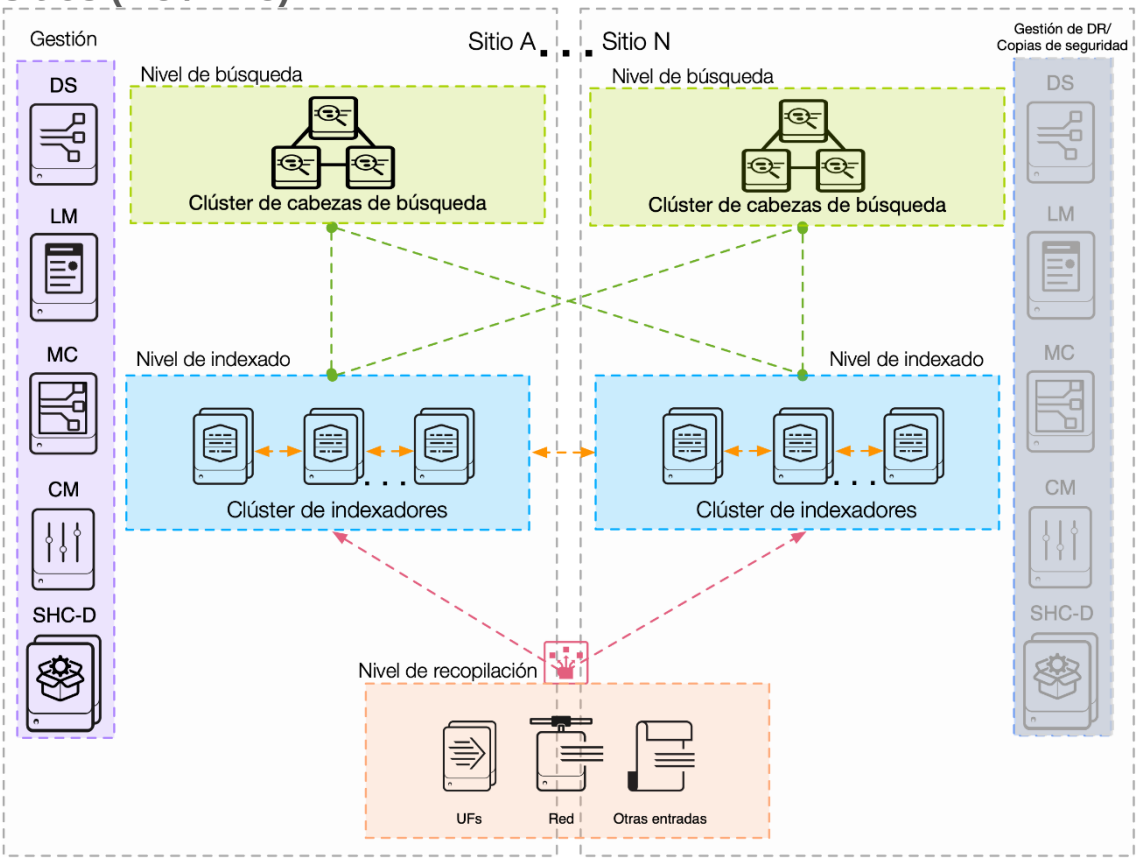
Descripción de la implementación agrupada en clúster distribuida + SHC - Único sitio (C3 / C13)	Limitaciones
La SHC proporciona el mecanismo para aumentar la capacidad de búsqueda disponible más allá de lo que una única cabeza de búsqueda puede proporcionar. Del mismo modo, la SHC permite la distribución de la carga de trabajo de búsqueda programada en todo el clúster. La SHC proporciona también una recuperación de fallos de usuario óptima en caso del fallo de una cabeza de búsqueda. Se requiere un equilibrador de carga de red que admita afinidad de sesiones delante de los miembros de la SHC para garantizar el equilibrio adecuado de la carga de usuarios en todo el clúster. **Nota para clientes de ES:** Si su código de categoría es C13 (por ej. tiene intención de implementar la Aplicación Splunk para la seguridad empresarial), se requiere un clúster de cabezas de búsqueda **en exclusiva** para implementar la aplicación (esto no se ilustra en el diagrama de topologías). El nivel de búsqueda puede contener cabezas de búsqueda en clústeres o no dependiendo de su capacidad y necesidades organizativas (esto no se ilustra en el diagrama de topologías).	• La SHC no puede tener más de 100 nodos

Implementación agrupada en clúster distribuida - Varios sitios (M2 / M12)



Descripción de la implementación agrupada en clúster distribuida - Varios sitios (M2 / M12)	Limitaciones
Para proporcionar recuperación de desastres casi automática en caso de un suceso catastrófico (como la interrupción de un centro de datos), la agrupación en clústeres de varios sitios es la arquitectura de implementación que hay que elegir. Una agrupación en clúster de varios sitios en buen estado requiere una latencia de red entre sitios aceptable según se especifica en la documentación de Splunk. Esta topología le permite replicar datos **de manera determinista** a dos o más grupos de puntos de clústeres de indexadores. Podrá configurar la replicación del sitio y el factor de búsqueda. Este factor de replicación del sitio le permite especificar la ubicación a la que se están enviando las copias de réplica y garantiza que los datos se distribuyen entre varias ubicaciones. Esto aún lo gestiona un único nodo maestro de clúster, que tiene que tener recuperación de fallos en el sitio de recuperación de desastres en caso de fallo. La agrupación en clústeres de varios sitios proporciona redundancia de datos entre ubicaciones distribuidas separadas físicamente, con la posibilidad de una distribución separada geográficamente. Los usuarios pueden recuperar errores en el sitio de recuperación de desastres automáticamente para garantizar la disponibilidad. No obstante, esta topología no proporciona un mecanismo para sincronizar automáticamente la configuración y los artefactos en tiempo de ejecución del nivel de búsqueda entre sitios. Se puede utilizar la capacidad del punto de búsqueda (indexador) disponible entre sitios para la ejecución de búsquedas en un modelo activo/activo. Cuando sea posible, la afinidad de los sitios puede configurarse para garantizar que los usuarios que se conectaron a una cabeza de búsqueda de un sitio específico solo buscarán en indexadores locales. **Nota para clientes de ES:** Si su código de categoría es M12 (por ej. tiene intención de implementar la Aplicación Splunk para la seguridad empresarial), se requiere una cabeza de búsqueda **única en exclusiva** para implementar la aplicación (esto no se ilustra en el diagrama de topologías). Para la cabeza de búsqueda de ES, la recuperación de fallos conlleva el establecimiento de una cabeza de búsqueda de "sombra" en el sitio de recuperación de fallos que solo se activa y se utiliza en una situación de recuperación de desastres. Participe con los Servicios Profesionales de Splunk para diseñar e implementar un mecanismo de recuperación de fallos de sitio para su implementación de Seguridad empresarial.	• No se puede compartir la capacidad de la cabeza de búsqueda disponible ni tampoco replicación de artefactos de búsqueda **entre sitios** • Las funciones de fallo de gestión tienen que tratarse fuera de Splunk en caso de un fallo de sitio • La latencia entre sitios para la replicación de índices debe encontrarse dentro de los límites recomendados

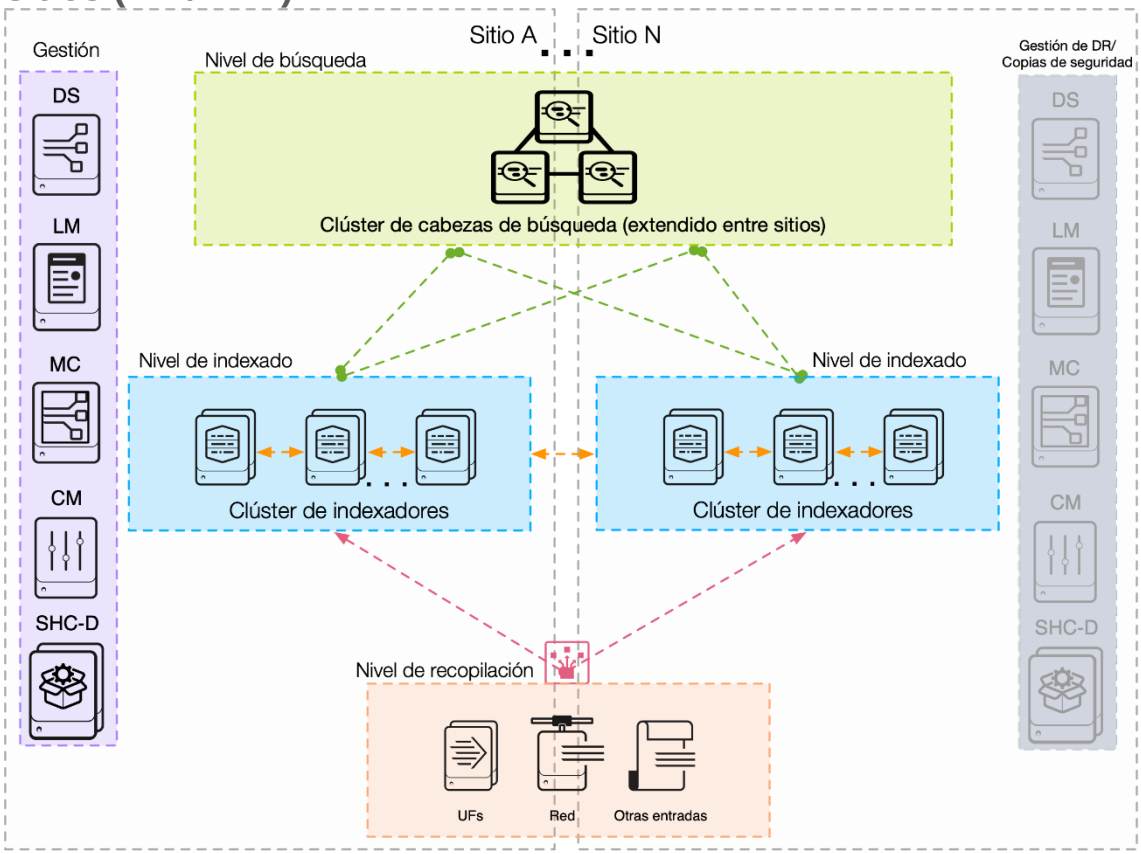
Implementación agrupada en clúster distribuida + SHC - Varios sitios (M3 / M13)



Descripción de la implementación agrupada en clúster distribuida + SHC - Varios sitios (M3 / M13)	Limitaciones
Esta topología añade capacidad de ampliación horizontal y elimina el punto de fallo único del nivel de búsqueda en cada sitio. Se requiere un mínimo de tres cabezas de búsqueda para implementar una SHC (por sitio). Para gestionar la configuración de la SHC, se requiere un componente de Splunk adicional denominado Implementador de clúster de cabezas de búsqueda **para cada** SHC. Este componente es necesario para implementar cambios en los archivos de configuración del clúster. El Implementador de clúster de cabezas de búsqueda no tiene requisitos de alta disponibilidad (sin función en tiempo de ejecución). La SHC proporciona los siguientes beneficios: a) capacidad de búsqueda disponible aumentada más allá de lo que una única cabeza de búsqueda puede proporcionar; b) distribución de la carga de trabajo de búsquedas programada entre el clúster; c) recuperación de fallos de usuario óptima en caso del fallo de una cabeza de búsqueda. Se requiere un equilibrador de carga de red que admita afinidad de sesiones delante de los miembros de la SHC en cada sitio para garantizar el equilibrio adecuado de la carga de usuarios en todo el clúster. **Nota para clientes de ES:** Si su código de categoría es M13 (por ej. tiene intención de implementar la Aplicación Splunk para	• No hay replicación de artefactos de búsqueda entre sitios, las SHC son autónomas • La latencia entre sitios para la replicación de índices debe encontrarse dentro de los límites documentados • La SHC no puede tener más de 100 nodos

Descripción de la implementación agrupada en clúster distribuida + SHC - Varios sitios (M3 / M13)	Limitaciones
la seguridad empresarial), se requiere un único clúster de cabezas de búsqueda en exclusiva *dentro de un sitio* para implementar la aplicación (esto no se ilustra explícitamente en el diagrama de topologías). Para poder recuperar un entorno SH de ES de un fallo de sitio, puede utilizarse tecnología externa para realizar una recuperación de fallos de las instancias de las cabezas de búsqueda, o puede aprovisionarse una "espera en caliente" SH de ES y mantenerse en sincronía con el entorno de ES principal. Se recomienda encarecidamente participar con los Servicios Profesionales de Splunk al implementar ES en un entorno de alta disponibilidad/recuperación de desastres.	

Implementación agrupada en clúster distribuida + SHC - Varios sitios (M4 / M14)



Descripción de la implementación agrupada en clúster distribuida + SHC - Varios sitios (M4 / M14)	Limitaciones
Esta es la arquitectura validada más compleja, diseñada para implementaciones que tienen requisitos estrictos en lo referente a la alta disponibilidad y la recuperación de desastres. Recomendamos encarecidamente que los Servicios Profesionales de Splunk participen para una implementación adecuada. Cuando se realiza una implementación adecuada, esta topología proporciona un funcionamiento continuado de su infraestructura de Splunk para la recopilación, el indexado y la búsqueda de datos.	• La latencia de red entre sitios debe estar dentro de los límites documentados • La recuperación de fallos de la SHC podría requerir pasos manuales si solo sobrevive una minoría de los miembros del clúster

Descripción de la implementación agrupada en clúster distribuida + SHC - Varios sitios (M4 / M14)	Limitaciones
Esta topología conlleva la implementación de un clúster de cabezas de búsqueda "extendido" que abarca uno o más sitios. Esto proporciona una recuperación de fallos óptima para los usuarios en caso de un fallo de nodo de búsqueda o centro de datos. Algunos artefactos y otros objetos de conocimiento en tiempo de ejecución se replican en la SHC. Se requiere una cuidadosa configuración para garantizar que la replicación se producirá entre los sitios, ya que la SHC en sí no detecta los sitios (por ej. la replicación de los artefactos no es determinista). Puede configurarse la afinidad de los sitios para garantizar que el vínculo WAN entre sitios se utiliza únicamente en los casos en los que una búsqueda no puede realizarse localmente. Se requiere un equilibrador de carga de red que admita afinidad de sesiones delante de los miembros de la SHC para garantizar el equilibrio adecuado de la carga de usuarios en todo el clúster. **Nota para clientes de ES:** Si su código de categoría es M14 (por ej. tiene intención de implementar la Aplicación Splunk para la seguridad empresarial), se requiere un único clúster de cabezas de búsqueda en exclusiva *dentro de un sitio* para implementar la aplicación (esto no se ilustra explícitamente en el diagrama de topologías). ES requiere que haya disponible un conjunto coherente de artefactos en tiempo de ejecución y esto no puede garantizarse en una SHC extendida cuando se produce la interrupción de un sitio. Para poder recuperar un entorno SH de ES de un fallo de sitio, puede utilizarse tecnología externa para realizar una recuperación de fallos de las instancias de las cabezas de búsqueda, o puede aprovisionarse una "espera en caliente" SH de ES y mantenerse en sincronía con el entorno de ES principal. Se recomienda encarecidamente participar con los Servicios Profesionales de Splunk al implementar ES en un entorno de alta disponibilidad/recuperación de desastres.	

Paso 1b: Definir sus requisitos de recopilación de datos

Su nivel de recopilación de datos es un componente principal de una implementación de Splunk. Permite a cualquier dispositivo de su entorno reenviar datos al nivel de indexado para su procesamiento, haciendo por ello que queden disponibles para su búsqueda en Splunk. El factor más importante aquí es garantizar que el reenvío y la indexación se produzcan de la manera más eficiente y fiable, ya que esto es vital para el éxito y el rendimiento de su implementación de Splunk.

Considere los siguientes aspectos de su arquitectura de nivel de recopilación de datos:

- El origen de sus datos. ¿Provienen de archivos de registros, fuentes syslog, entradas de red, instalaciones de registro de eventos de SO, aplicaciones, bus de mensajes o de otros lugares?
- Requisitos para la latencia y el rendimiento de la introducción de datos
- Distribución de eventos ideal entre los indexadores en su nivel de indexado
- Tolerancia a fallos y recuperación automática (alta disponibilidad)

- Requisitos de soberanía de seguridad y datos

Esta sección de las SVA se centra en los métodos de recopilación de datos comunes. Esta sección también trata la arquitectura y las prácticas recomendadas para cada método de recopilación de datos y enumera posibles problemas a tener en cuenta al realizar su elección de implementación.

Consideraciones arquitectónicas importantes y por qué son relevantes

Dada la función esencial del nivel de recopilación de datos, es importante comprender las consideraciones clave que conlleva el diseño de la arquitectura.

Aunque algunas de estas consideraciones pueden ser o no relevantes para usted en base a sus requisitos, las consideraciones en texto en negrita de la tabla que aparece a continuación describen elementos fundamentales que son relevantes para cada entorno.

Consideración	**¿Por qué es importante?**
Los datos se introducen adecuadamente (marcas de tiempo, saltos de línea, truncamiento)	La importancia de la distribución de eventos ideal entre indexadores no puede sobrevalorarse. El nivel de indexado funciona de la manera más eficiente cuando todos los indexadores disponibles se utilizan de forma igualitaria. Esto es así tanto para la introducción de los datos como para el rendimiento de las búsquedas. Un único indexador que controle significativamente más introducciones de datos en comparación con otros puntos puede afectar negativamente a los tiempos de respuesta de las búsquedas. Para los indexadores con un almacenamiento en disco local limitado, la distribución de eventos desigual también podría provocar el envejecimiento prematuro de los datos antes de cumplir la directriz de retención de datos configurada.
Los datos se distribuyen de manera óptima entre los indexadores disponibles	Si los datos no se introducen adecuadamente debido a una mala configuración de las marcas de tiempo de los eventos y los saltos de línea, la búsqueda de estos datos se volverá muy difícil. Esto se debe a que los límites de los eventos deben respetarse en el momento de la búsqueda. Las configuraciones de extracción de marcas de tiempo incorrectas o inexistentes pueden producir una asignación de marcas de tiempo implícitas no deseada. Esto confundirá a sus usuarios y la obtención de valor de sus datos será mucho más difícil de lo necesario.
Todos los datos alcanzan el nivel de indexado de forma fiable y sin pérdidas	Todos los datos de registro que se recopilen para el propósito de realizar análisis fiables tienen que ser completos y válidos, de modo que las búsquedas realizadas en los datos proporcionen datos válidos y precisos.
Todos los datos alcanzan el nivel de indexado con la mínima latencia	Las demoras en la introducción de los datos aumentará el tiempo entre la aparición de un evento potencialmente crítico y la capacidad de buscar y reaccionar ante él. La latencia de introducción mínima es a menudo crucial para supervisar casos de uso que desencadenan alertas para la plantilla o provocan una acción automatizada.
Los datos se aseguran en el tránsito	Si los datos son confidenciales o es necesario protegerlos mientras se envían por redes no de confianza, es posible que se requiera el cifrado de los datos para evitar la intercepción no autorizada de terceros. Normalmente, recomendamos que todas

	las conexiones entre componentes de Splunk se realicen con SSL activado.
El uso de recursos de red se reduce al mínimo	La repercusión en los recursos de red de la recopilación de datos de registro debe reducirse al mínimo para que no afecte a otro tráfico de red vital para la actividad comercial. Para redes con líneas arrendadas, la reducción al mínimo de la utilización de la red también contribuye a un coste total de propiedad inferior de su implementación.
Autenticar/autorizar fuentes de datos	Para evitar que fuentes de datos malintencionadas afecten a su entorno de indexado, considere la implementación de una autenticación/autorización de conexiones. Esto puede realizarse empleando controles de red, o bien empleando mecanismos en el nivel de aplicaciones (por ej. SSL/TLS).

Debido a su función vital en su implementación, la orientación en este documento se centra en arquitecturas que fomentan la distribución ideal de los eventos. Cuando un entorno de Splunk no proporciona el rendimiento de las búsquedas esperado, en la mayoría de los casos viene provocado por no cumplir los requisitos de rendimiento mínimos del almacenamiento y/o por una distribución desigual de los eventos que limitan el aprovechamiento de la paralelización de las búsquedas.

Ahora que comprende las consideraciones arquitectónicas más críticas, averigüemos qué requisitos de recopilación de datos específicos necesita satisfacer.

Cuestionario 2: Definición de sus requisitos de recopilación de datos

La respuesta a las siguientes preguntas le darán una lista de los componentes de recopilación de datos que necesita en su implementación. Puede utilizar las claves de la columna más a la derecha para encontrar más detalles sobre cada componente más adelante en el documento.

Nº	Pregunta	Consideraciones	Repercusión sobre la topología	Componentes de recopilación de datos relevantes
1	¿Necesita supervisar archivos locales o ejecutar secuencias de recopilación de datos en extremos?	Este es un requisito principal para todos los escenarios de implementación de Splunk.	Tendrá que instalar el reenviador universal en sus extremos y gestionar su configuración centralmente.	UF
2	¿Necesita recopilar datos de registro enviados a través de syslog desde dispositivos donde no puede instalar software (aparatos, conmutadores de red, etc.)?	Syslog es un protocolo de transporte ubicuo utilizado a menudo por dispositivos construidos para un propósito específico que no permiten la instalación de software personalizado.	Necesitará una infraestructura de servidores syslog que sirva como el punto de recopilación.	SYSLOG HEC
3	¿Necesita admitir la recopilación de datos de registro desde aplicaciones	La escritura de archivos de registro en extremos requiere proporcionar espacio en disco y la gestión de estos archivos	Tendrá que utilizar el Recopilador de eventos HTTP de Splunk (HEC) u otra	HEC

	que registran en una API frente a la escritura en discos locales?	de registro (rotación, eliminación, etc.). Algunos clientes desean apartarse de este modelo y registrar directamente en Splunk empleando las bibliotecas de registro disponibles.	tecnología que sirva como receptor de registro.	
4	¿Necesita recopilar datos desde un proveedor de datos de eventos de transmisión?	Muchas empresas han adoptado un modelo de concentración de eventos donde una plataforma de datos de transmisión centralizada (como Kinesis de AWS o Kafka) sirve como el transporte de mensajes entre los productores de datos de registro y los consumidores.	Necesitará una integración entre el proveedor de datos de transmisión y Splunk.	KAFKA KINESIS HEC
5	¿Tiene directrices de seguridad no negociables que evitan que los productores de datos de registro establezcan conexiones TCP directamente con el nivel de indexado?	A veces, las topologías de red constan de varias zonas de red con reglas de cortafuegos restrictivas entre ellas y puede que no sea posible permitir genéricamente que fluya entre zonas el tráfico en puertos de Splunk. La configuración y el mantenimiento de reglas de cortafuegos para direcciones IP de origen/destino podría ser demasiado complejo.	Necesitará un nivel de reenvío intermedio que permita que fluya el tráfico entre zonas de red.	IF
6	¿Necesita recopilar datos de registro empleando medios programáticos, por ej. llamando a las API de REST o consultando bases de datos?	Splunk proporciona varias entradas modulares que permiten la ejecución de secuencias en las API para obtener una gran variedad de casos de uso de introducción de datos, incluyendo DBX para la recopilación de datos procedentes de bases de datos relacionales.	Su nivel de recopilación de datos requerirá uno o más nodos de recopilación de datos (DCN) implementados con un Reenviador de alta intensidad de Splunk.	DCN
7	¿Necesita enrutar (un subconjunto de) datos a otros sistemas aparte de (y como adición a) Splunk?	Algunos casos de uso requieren que los datos que se indexan en Splunk también se reenvíen a otro sistema. A menudo, los datos reenviados constan únicamente de un subconjunto de los datos de origen, o los datos tienen que modificarse antes de reenviarse.	Dependiendo de los aspectos específicos del caso de uso, es posible que necesite un nivel de reenvío intermedio construido equipado con un Reenviador de alta intensidad para dar cobertura al enrutamiento y filtrado basado en eventos. De manera alternativa, puede reenviar datos posteriormente a la indexación utilizando el comando cefout	HF

			contenido en la Aplicación Splunk para CEF.	
8	¿Tiene sitios remotos con restricciones de ancho de banda de red y requiere un filtrado significativo de los datos antes de que se envíen por la red?	El filtrado de los datos antes de la transmisión requiere un reenviador de análisis (de alta intensidad). El ancho de banda de red saliente empleado por un reenviador de alta intensidad es aproximadamente cinco veces el que utiliza un reenviador universal, de modo que el filtrado solo tiene sentido si se excluye con el filtrado un número significativo de eventos (norma general: >50% de los datos de origen). Idealmente, debería ajustar la granularidad de su operación de registro para alcanzar la reducción necesaria en el volumen del registro.	Si no puede reducir su volumen del registro en el origen, necesitará un reenviador de alta intensidad intermedio en su sitio remoto que analice los datos de origen y excluya filtrando los eventos basados en la configuración.	IF HF
9	¿Necesita enmascarar/ofuscar datos confidenciales antes de que se envíen por una red pública para su indexado?	A veces, asegurar el tráfico de reenviadores con SSL no es suficiente para proteger los datos confidenciales en tránsito por redes públicas y deben enmascararse partes de eventos individuales antes de la transmisión (SSN, datos CC, etc.). De manera ideal, dicho enmascarado de datos se realizará en la aplicación que genera los datos de registro.	Si no puede enmascarar los datos en la aplicación generadora, necesitará un reenviador de alta intensidad intermedio en su sitio que analice los datos de origen y aplique las reglas de enmascaramiento requeridas en base a la configuración antes de que se envíen los datos a los indexadores.	IF HF
10	¿Necesita capturar mediciones empleando statsd o collectd?	Statsd y collectd son tecnologías ubicuas que se utilizan para recopilar mediciones procedentes de sistemas y aplicaciones anfitriones.	Splunk admite tipos de índices y métodos de recopilación específicos para transmitir esos índices empleando UF, HF o HEC.	MEDICIONES
11	¿Necesita que cualquiera de sus componentes de recopilación de datos tenga alta disponibilidad?	La disponibilidad, que normalmente no se aplica a los extremos, puede ser un problema para otros componentes de recopilación de datos, como reenviadores intermedios o nodos de recopilación de datos.	Se necesitan perspectivas sobre cómo las interrupciones afectarán a la disponibilidad de cada componente, y cómo solucionar esto.	HA

Paso 2b: Seleccionar sus componentes de recopilación de datos

Después de completar el cuestionario, tendrá una lista de los componentes de recopilación de datos requeridos para cumplir su implementación. Esta sección trata cada componente de la arquitectura de recopilación de datos con mayor detalle. Antes de hacerlo, proporcionemos brevemente algunas directrices generales.

Directrices generales sobre la arquitectura de reenvío

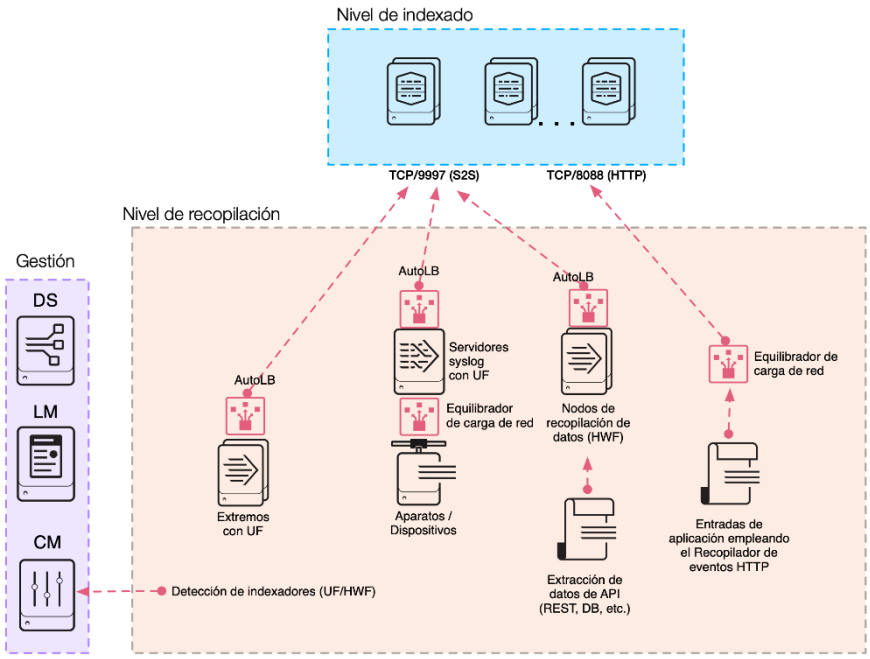
idealmente, el nivel de recopilación de datos es tan "plano" como sea posible. Esto significa que las fuentes de datos se recopilan localmente por parte de un reenviador universal y reenviados directamente al nivel de indexado. Esto es una práctica recomendada porque garantiza la latencia mínima en la introducción de los datos (el tiempo de búsqueda) y permite una distribución adecuada de los eventos entre los indexadores disponibles. El seguimiento de esta práctica recomendada lleva a una facilidad de la gestión y la sencillez operativa. A menudo vemos clientes implementar un nivel de reenvío intermedio. Por lo general, evite esto a no ser que no se puedan cumplir los requisitos de otra manera. Debido a la potencial repercusión de los reenviadores intermedios, este documento contiene una sección separada sobre este tema con más detalles.

Existen extremos que no permiten la instalación del reenviador universal (en otras palabras, dispositivos de red, aparatos) y registran empleando el protocolo syslog. Una arquitectura de práctica recomendada separada para recopilar dichas fuentes de datos se describe en la sección Recopilación de datos syslog.

Para las fuentes de datos que hay que recopilar empleando medios programáticos (API, acceso de base de datos), se recomienda la implementación de un nodo de recopilación de datos (DCN) basado en una instalación completa de Splunk Enterprise. Esto también se conoce como reenviador de alta intensidad. No se recomienda que ejecute estos tipos de entradas en el nivel de cabezas de búsqueda en otra cosa que no sea un entorno de desarrollo.

Los siguientes diagramas muestran una arquitectura de recopilación de datos general que refleja estas directrices.

Descripción general de la topología de recopilación de datos



El diagrama anterior muestra el Servidor de implementación (DS) en el nivel de gestión, que se utiliza para gestionar las configuraciones en los componentes de recopilación de datos. Del mismo modo, el Maestro de licencias (LM) se muestra aquí, ya que los nodos de recopilación de datos requieren acceso al LM para activar las funciones de Splunk Enterprise. Los reenviadores pueden utilizar el maestro de clúster (CM), si está disponible, para la detección de indexadores, eliminando la necesidad de gestionar los indexadores disponibles en la configuración de resultados del reenviador.

En el diagrama anterior, AutoLB representa el mecanismo de equilibrio de cargas automático incorporado en Splunk. Este mecanismo se utiliza para garantizar la distribución adecuada de los eventos para los datos enviados empleando el protocolo S2S creado por Splunk (puerto predeterminado 9997). Nota: El uso de un equilibrador de carga de red para el tráfico S2S no se admite en estos momentos y no se recomienda.

Para equilibrar la carga del tráfico procedente de fuentes de datos que se comunican con un protocolo estándar del sector (como HTTP o syslog), se utiliza un equilibrador de carga de red para garantizar cargas y distribución de eventos uniformes entre todos los indexadores del nivel de indexado.

Reenviador universal (UF)

El reenviador universal (UF) es la mejor opción para requisitos de recopilación de datos de conjuntos de gran tamaño procedentes de sistemas de su entorno. Es un mecanismo de recopilación de datos construido específicamente con unos requisitos de recursos reducidos al mínimo. El UF debería ser la opción predeterminada para la recopilación y reenvío de datos de registro. El UF proporciona:

- Función de punto de control/reinicio para una recopilación de datos sin pérdidas.
- Protocolo eficiente que reduce al mínimo la utilización del ancho de banda de red.
- Capacidades de aceleración.
- Equilibrio de carga incorporado entre los indexadores disponibles.
- Cifrado de red opcional empleando SSL/TLS.
- Compresión de datos (usar únicamente sin SSL/TLS)
- Múltiples métodos de entrada (archivos, registros de eventos de Windows, entradas de red, entradas de secuencias).
- Funciones de filtrado de eventos limitadas (solo registros de eventos de Windows).
- Compatibilidad con un canal de introducción paralelo para aumentar el rendimiento y reducir la latencia.

Con algunas excepciones para los datos bien estructurados (json, csv, tsv), el UF no analiza las fuentes de registro en eventos, de modo que no puede realizar ninguna acción que requiera la comprensión del formato de los registros. También se entrega con una versión reducida de Python, lo que lo hace incompatible con cualquier aplicación de entrada modular que requiera una pila de Splunk completa para funcionar.

Es normal que se implemente un gran número de UF (de 100 a 10.000) en extremos y servidores en un entorno de Splunk y que se gestionen de manera centralizada, ya sea con un servidor de implementación de Splunk o con una herramienta de gestión de configuración externa (como por ej. Puppet o Chef).

Reenviador de alta intensidad (HF)

El reenviador de alta intensidad (HWF) es una implementación completa de Splunk Enterprise configurada para actuar como reenviador con el indexado desactivado. Un HWF normalmente no realiza otras funciones de Splunk. La diferencia clave entre un UF y un HWF es que el HWF contiene el canal de análisis completo, y realiza funciones idénticas a las de un indexador, sin escribir e indexar realmente eventos en disco. Esto permite que el HWF comprenda y actúe sobre eventos individuales, como por ejemplo para enmascarar datos o realizar filtrado y

enrutamiento basado en datos de los eventos. Ya que es una instalación completa de Splunk Enterprise, puede alojar entradas modulares que requieran una pila de Python completa para funcionar adecuadamente para la recopilación de datos o servir como un extremo para el Recopilador de eventos HTTP de Splunk (HEC). El HWF realiza las funciones siguientes:

- Analiza los datos en eventos.
- Filtra y enruta en base a datos de eventos individuales.
- Tiene una presencia de recursos más grande que el UF.
- Tiene una presencia de ancho de banda de red más grande que el UF (~5x).
- Interfaz gráfica de usuario para la gestión.

En general, los HWF no se instalan en extremos para el propósito de recopilación de datos. En su lugar, se utilizan en sistema autónomos para implementar nodos de recopilación de datos (DCN) o niveles de reenvío intermedios. **Utilice un HWF únicamente cuando los requisitos para recopilar datos de otros sistemas no se pueda realizar con un UF.**

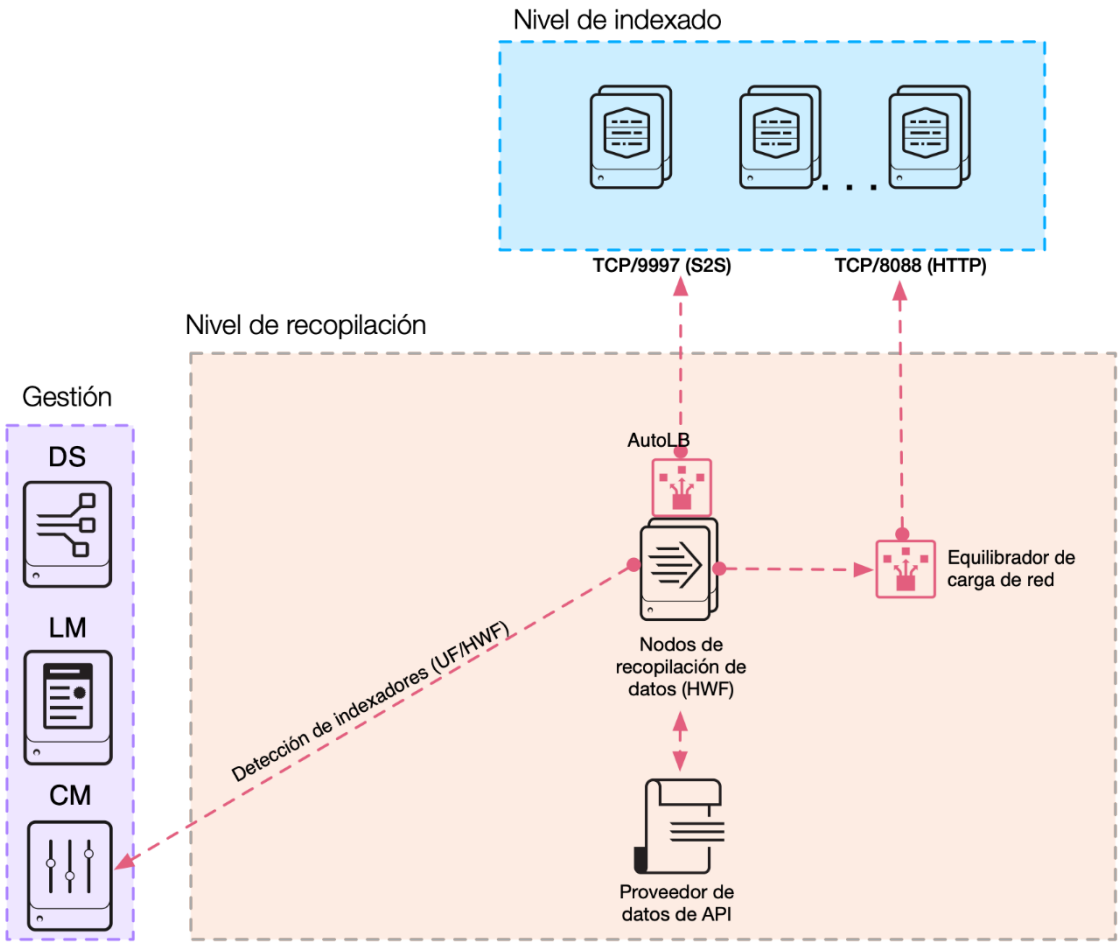
Los ejemplos de dichos requisitos incluyen:

- Lectura de datos procedentes de RDBMS para el propósito de introducirlos en Splunk (entradas de base de datos).
- Recopilación de datos de sistemas a los que se accede a través de una API (servicios de nube, supervisión VMWare, sistemas propios, etc.).
- Proporcionar un nivel dedicado para alojar el servicio recopilador de eventos HTTP.
- La implementación de un nivel de reenvío intermedio requiere un reenviador de análisis para el enrutamiento/filtrado/enmascarado.

Reenviador de alta intensidad (DCN) como nodo de recopilación de datos

Algunas fuentes de datos requieren la recopilación empleando algún tipo de API. Estas API pueden incluir REST, servicios web, JMS y/o JDBC como mecanismo de consulta. Splunk, así como otros desarrolladores externos, proporcionan una amplia variedad de aplicaciones que permiten que se produzcan esas interacciones de API. Más comúnmente, estas aplicaciones se implementan utilizando el marco de trabajo Entrada modular de Splunk, que requiere una instalación completa del software Splunk Enterprise para funcionar adecuadamente. La práctica recomendada para realizar este caso de uso es implementar uno o más servidores para que funcionen como un reenviador de alta intensidad configurado para funcionar como un Nodo de recopilación de datos (DCN).

Topología de nodos de recopilación de datos

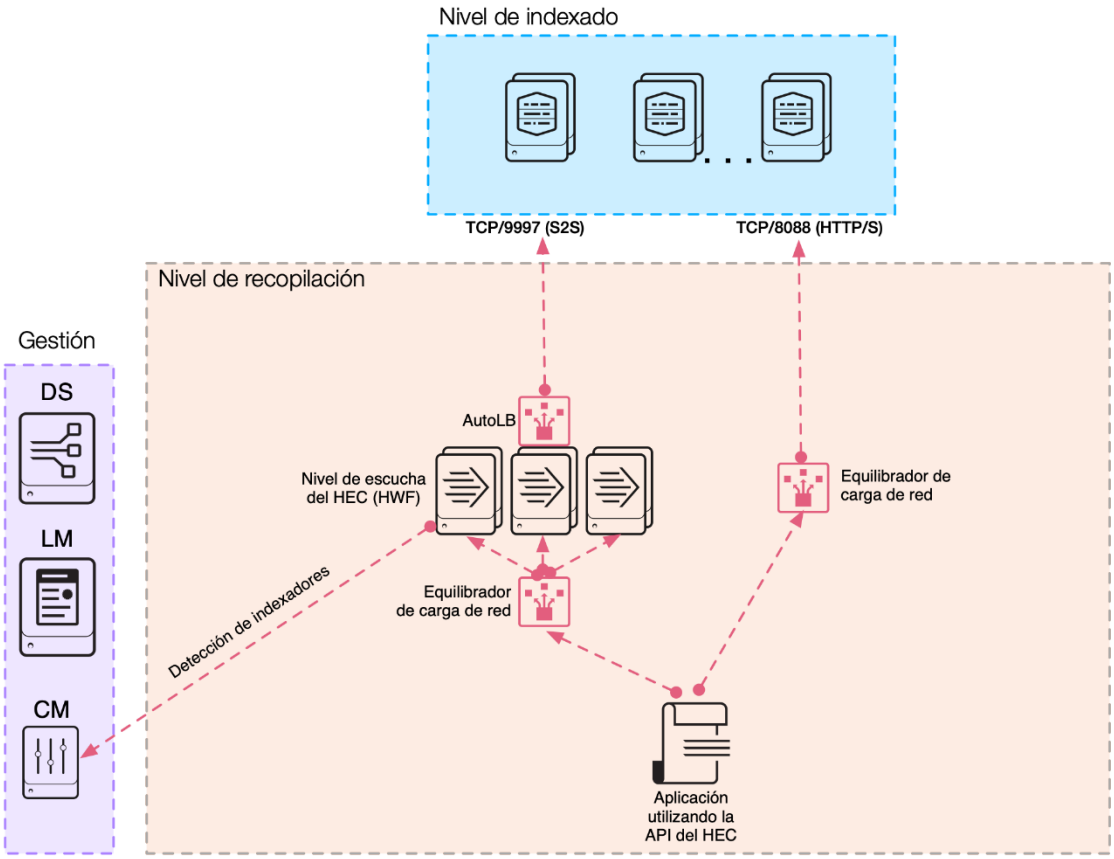


Recopilador de eventos HTTP (HEC)

El HEC proporciona un servicio de escucha que acepta conexiones HTTP/S en el lado del servidor, así como una API en el lado del cliente, lo que permite que las aplicaciones publiquen cargas de trabajo de datos de registro directamente en el nivel de indexado o un nivel de receptor HEC dedicado que conste de uno o más reenviadores de alta intensidad. El HEC proporciona dos extremos que admiten el envío de los datos en un formato sin procesar o en formato JSON. El uso de JSON puede permitir que se incluyan metadatos adicionales en la carga de trabajo de eventos que puede facilitar una flexibilidad mayor cuando se busca en los datos más adelante.

El diagrama siguiente ilustra las dos opciones de implementación para el HEC:

Elecciones de topología del HEC



El nivel de gestión contiene el Maestro de licencias (requerido por el HF) así como el servidor de implementación para gestionar las entradas HTTP en los componentes de escucha. Nota: Si el nivel de indexado está agrupado en clústeres y recibe tráfico del HEC directamente, la configuración del HEC se gestiona a través del maestro de clúster en vez del servidor de implementación.

La decisión de qué topología de implementación elija depende en gran medida de sus necesidades específicas. Un nivel de escucha de HEC dedicado introduce otro componente arquitectónico en su implementación. En el lado positivo, puede ampliarse de forma independiente y proporciona un nivel de aislamiento del nivel de indexado desde una perspectiva de gestión. Del mismo modo, ya que el nivel de HEC dedicado requiere un HF, analizará todo el tráfico entrante, retirando esa carga de trabajo de los indexadores.

Por otra parte, alojar la escucha del HEC directamente en los indexadores probablemente garantizará una mejor distribución de los eventos en el nivel de indexado, debido a que HTTP es un protocolo bien comprendido por todos los equilibradores de carga de red y la directriz de equilibrio de carga apropiada puede ayudar a garantizar que los indexadores menos ocupados reciban la carga primero.

Siguiendo la idea de implementar la arquitectura más sencilla posible que cumpla sus requisitos, le recomendamos que considere alojar su escucha del HEC en los indexadores, asumiendo que dispone de la capacidad del sistema suficiente para ello. La decisión puede revertirse fácilmente más adelante si surge la necesidad sencillamente implementando un nivel de HF con la configuración y el tamaño apropiados y cambiando la configuración de carga de red para que utilice las direcciones IP del HF en vez de los indexadores. Ese cambio debería ser transparente para aplicaciones cliente.

Nota: Si no requiere que reconocimiento de indexadores para los datos enviados a través del HEC, se recomienda un nivel de escucha del HEC dedicado para reducir al mínimo los mensajes duplicados debido a reinicios graduales del indexador.

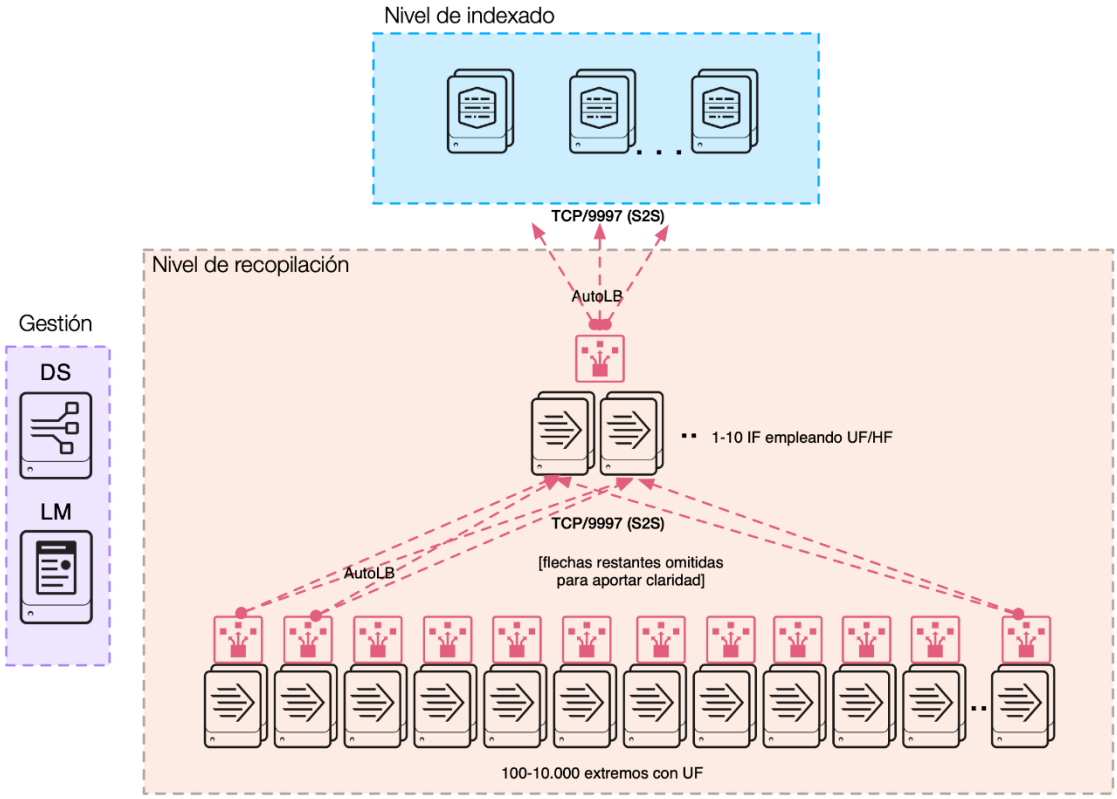
Nota: Esta arquitectura de implementación del HEC proporciona el transporte para algunos de los otros componentes de recopilación de datos tratados más adelante, específicamente la recopilación de datos de syslog y mediciones.

Nivel de reenvío intermedio (IF)

El algunas situaciones, los reenviadores intermedios se necesitan para el reenvío de los datos. Los reenviadores intermedios reciben transmisiones de registros procedentes de extremos y las reenvían a un nivel de indexado. Los reenviadores intermedios introducen retos arquitectónicos que requieren un diseño cuidadoso para evitar repercusiones negativas sobre el entorno general de Splunk. Como aspecto más destacado, los reenviadores intermedios concentran las conexiones procedentes de 100 a 10.000 reenviadores de extremo y reenvían a indexadores empleando un número muchísimo más reducido de conexiones. Esto puede afectar materialmente a la distribución de los datos en el nivel de indexado, ya que solo un subconjunto de indexadores están recibiendo tráfico en un punto específico en el tiempo. No obstante, estos efectos asociados negativos pueden mitigarse con un tamaño y una configuración adecuados.

El siguiente diagrama ilustra bien este reto:

Topología de reenvío intermedio



En un escenario con un único reenviador intermedio, todos los extremos se conectan con este único reenviador (potencialmente miles), y el reenviador intermedio a su vez solo se conecta con un indexador en algún momento dado. Este no es un escenario óptimo debido a que es posible que se produzcan las siguientes consecuencias:

- Una transmisión de datos de gran tamaño procedente de muchos extremos se canaliza a través de un único canal que agota su sistema y recursos de red.

- La recuperación de fallos limitada se dirige a los extremos en caso de fallo del IF (su riesgo de interrupción es inversamente proporcional al número de IF).
- Se proporciona carga a un número reducido de indexadores en algún momento dado. Las búsquedas durante periodos cortos de tiempo no se beneficiarán de la paralelización como podrían hacerlo de otro modo.

Los reenviadores intermedios también añaden un nivel arquitectónico adicional a su implementación, lo que puede complicar la gestión y la solución de errores y añade latencia a su ruta de introducción de datos. Intente evitar el uso de niveles de reenvío intermedios a no ser que sea la única opción para cumplir sus requisitos. Puede considerar utilizar un nivel intermedio si tiene:

- Datos confidenciales que necesitan ofuscarse/eliminarse antes del envío por la red a los indexadores. Un ejemplo es cuando debe utilizar una red pública.
- Las directrices de seguridad estrictas no permiten las conexiones directas entre extremos e indexadores, como las redes multizona o los indexadores basados en la nube.
- Las restricciones del ancho de banda entre extremos e indexadores requieren un subconjunto significativo de filtrado de eventos.
- El enrutamiento basado en eventos a objetivos dinámicos es un requisito.

Considere las necesidades de tamaño y configuración de cualquier nivel de reenvío intermedio para garantizar la disponibilidad de este nivel; proporcione capacidad de procesamiento suficiente para controlar todo el tráfico y fomente la buena distribución de los eventos entre indexadores. El nivel de IF tiene los siguientes requisitos:

- Número suficiente de canales de procesamiento de datos en general.
- Infraestructura de IF redundante.
- Una configuración de equilibrio de carga de Splunk adecuadamente ajustada. Por ejemplo, autoLBVolume, EVENT_BREAKER, EVENT_BREAKER_ENABLE, posiblemente forceTimeBasedAutoLB según se necesite.

La directriz general sugiere tener el doble de canales de procesamiento de IF que indexadores en el nivel de indexado.

Nota: Un canal de procesamiento no se equipara a un servidor de IF físico. Proporcione recursos del sistema suficientes. Por ejemplo, cuando hay disponibles núcleos de CPU, memoria y ancho de banda de NIC, un único IF puede configurarse con múltiples canales de procesamiento.

Si necesita un nivel de IF (consulte el cuestionario), tome de forma predeterminada el uso de UF para el nivel, ya que proporcionan un rendimiento superior con una presencia de recursos inferior tanto en el sistema como en la red. Utilice HF si su capacidad de IF no cumple sus requisitos.

Recopilación de datos syslog (SYSLOG)

El protocolo syslog ofrece una fuente ubicua de datos de registro en la empresa. Los niveles de recopilación de datos más ampliables y fiables contienen un componente de introducción de syslog. Hay muchas maneras de aportar datos de syslog a Splunk. Considere los métodos siguientes:

- **Reenviador universal (UF)/Reenviador de alta intensidad (HF)**: Utilice un UF o HF de Splunk para supervisar (introducir) archivos escritos por un servidor syslog (como `rsyslog` o `syslog-ng`).
- **Agente de syslog a HEC**: Utilice un agente de syslog que sea capaz de enviar resultados al HEC de Splunk. (Hay módulos externos para rsyslog y syslog-ng que pueden enviar resultados al HEC).

- **Entrada TCP/UDP directa**: Splunk tiene la capacidad de escuchar en un puerto TCP o UDP (el puerto predeterminado es UDP 514) e introducir fuentes aquí (**no** se recomienda para el uso en producción).

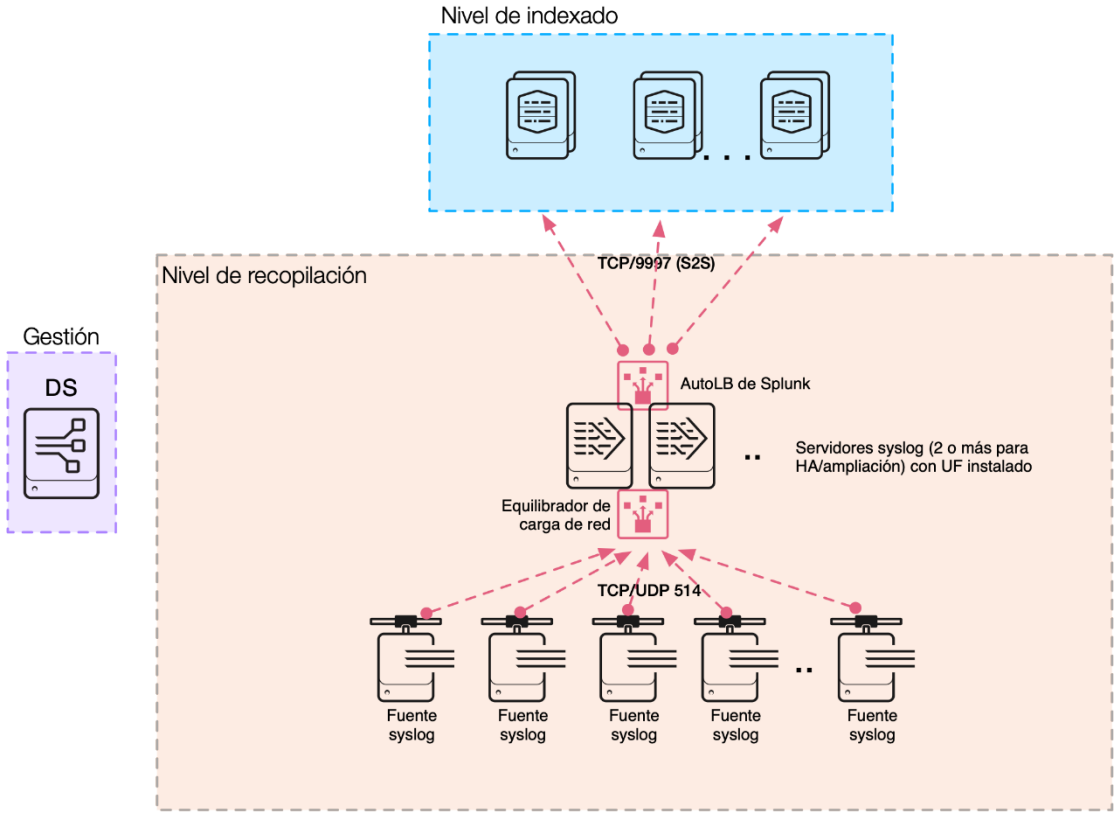
Syslog (supervisión de archivos en conjunción con un SCD)

Splunk puede utilizar la supervisión, con `inputs.conf`, en un UF/HF para procesar e introducir fuentes de syslog que se escriben en disco en un extremo a través de un centinela de recopilación de syslog (SCD). La opción que más se encuentra, tanto `rsyslog`, `syslog-ng` como Fastvue ofrecen soluciones comerciales y gratuitas que son ampliables y sencillas de integrar y gestionar en entornos de bajo volumen y entornos distribuidos a gran escala.

Para conocer más detalles sobre el modo de configurar los supervisores, consulte Monitor files and directories en *Getting Data In.*

Esta arquitectura admite la incorporación adecuada de los datos de la misma manera que lo hace un reenviador universal en cualquier otro extremo. Puede configurar el SCD para identificar múltiples tipos de registros diferentes y escribir los eventos de registro en los archivos y directorios apropiados donde un reenviador de Splunk los pueda recoger. Esto también añade un nivel de persistencia a la transmisión de registros de syslog escribiendo los eventos en disco, lo que puede limitar la exposición a la pérdida de datos para mensajes enviados empleando el UDP poco fiable como transporte.

Topología de recopilación de datos de syslog empleando UF



El diagrama muestra las fuentes syslog enviado datos utilizando TCP o UDP en el puerto 514 a un conjunto con equilibrio de carga de servidores syslog. Los múltiples servidores garantizan la alta disponibilidad para el nivel de recopilación y pueden evitar la pérdida de datos durante las operaciones de mantenimiento. Cada servidor syslog se configura para aplicar reglas a la transmisión de syslog que puede dar como resultado que los eventos de syslog se escriban en archivos/directorios dedicados para cada tipo de fuente (eventos de cortafuegos, syslog de SO, conmutadores de red, IPS, etc.) El UF que se implementa en cada servidor supervisa esos archivos y reenvía los datos al nivel de indexado para el procesamiento en el índice apropiado.

AutoLB de Splunk se utiliza para distribuir los datos de manera uniforme entre los indexadores disponibles.

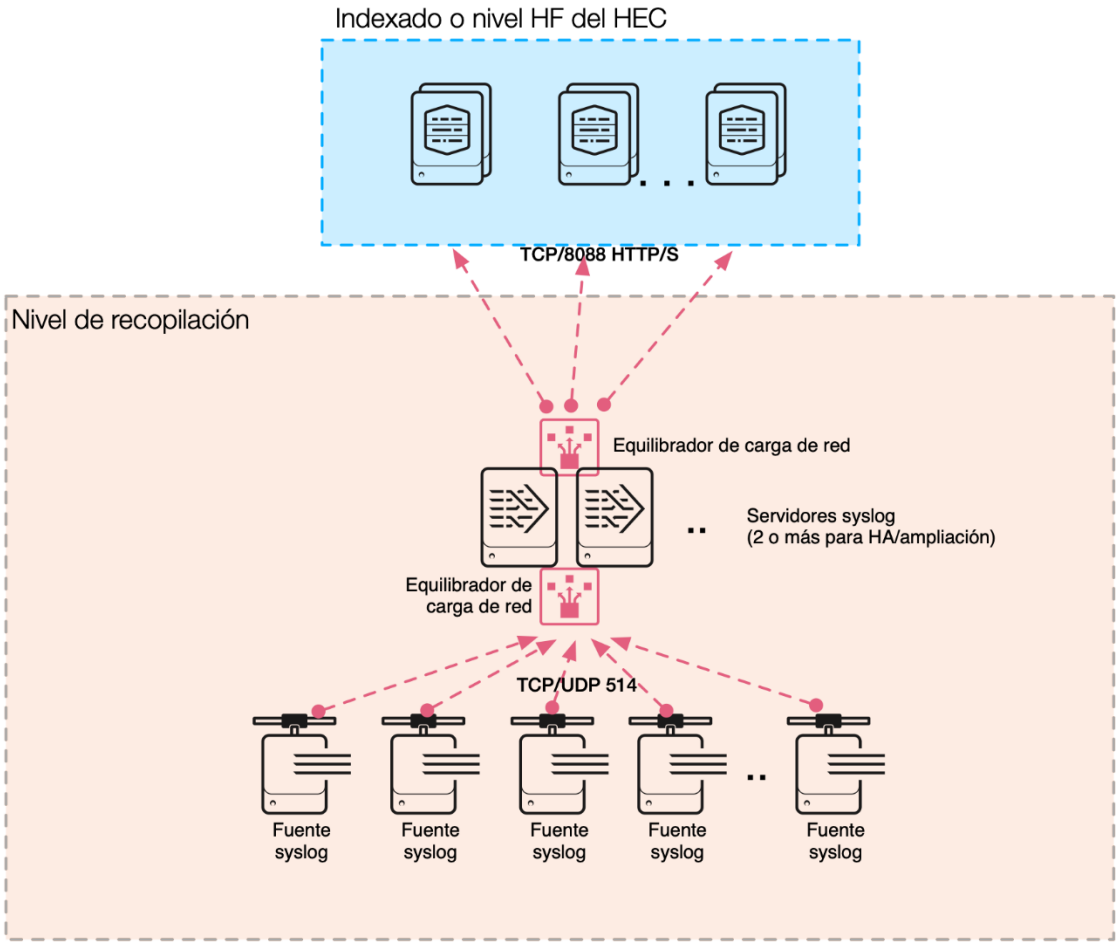
El servidor de implementación que se muestra en el nivel de gestión puede utilizarse para gestionar centralmente la configuración del UF.

Agente de syslog a HEC

Con la adopción aumentada del HEC, existe un número creciente de implementaciones que están utilizando sus implementaciones del HEC para introducir datos syslog. Para conocer más detalles, consulte la publicación del blog de Splunk Syslog-ng and HEC: Scalable Aggregated Data Collection in Splunk.

El siguiente diagrama muestra las fuentes syslog enviado datos en el puerto 514 utilizando un equilibrador de carga de red en una granja de servidores syslog. Se aplican las directrices de syslog apropiadas con un destino de syslog personalizado y una secuencia Python que emplea la API de HEC, y los eventos se envían a una escucha del HEC, junto con un equilibrador de carga de tráfico de red para la indexación:

Topología de recopilación de datos de syslog empleando HEC



Un beneficio de esta topología es que elimina la necesidad de implementar y configurar UF/HF. El equilibrador de carga HTTP sirve a las escuchas del HEC en los indexadores (o un nivel de escucha del HEC) para garantizar que los datos se están distribuyendo de manera uniforme entre los extremos del HEC. Configure este equilibrador de carga con la directriz "Mínimo de conexiones".

Entrada UDP de Splunk

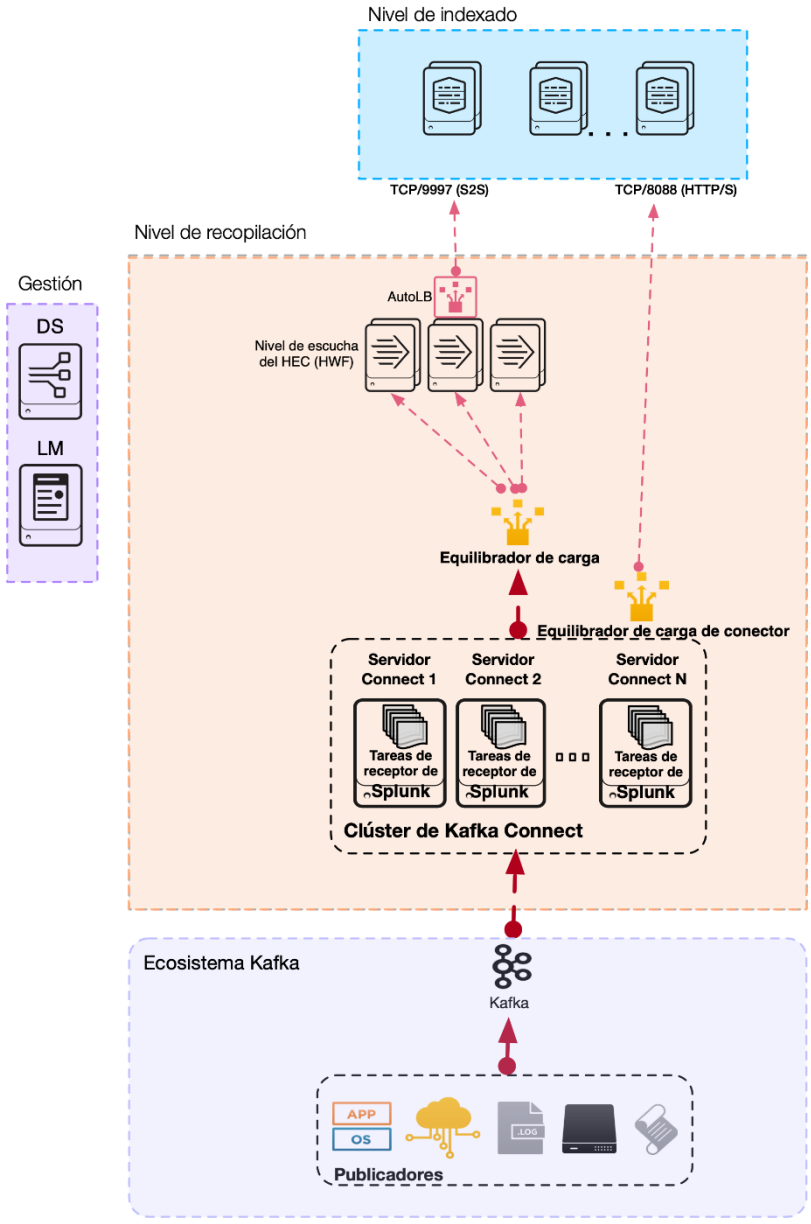
Splunk puede utilizar una entrada UDP directa en un UF o HF para recibir datos desde syslog. Para conocer más detalles sobre la configuración de los puertos TCP y UDP, consulte Get data

from TCP and UDP ports en *Getting Data In*. La capacidad de recibir eventos en el puerto 514 UDP se apoya en la capacidad del UF/HF de ejecutarse como raíz. De manera adicional, el agente debe estar disponible el 100% del tiempo para evitar una posible pérdida de datos. Los reenviadores pueden reiniciarse con frecuencia para aplicar cambios de configuración, lo que garantiza en gran medida una pérdida de datos. Por estas razones, **esta no se considera una práctica recomendada para la implementación en producción**.

(KAFKA) Consumo de datos de registro procedentes de temas de Kafka

Splunk proporciona un conector de receptor admitido para consumir datos procedentes de temas de Kafka denominado "Splunk Connect for Kafka". Consulte Apache Kafka Connect en el manual de Splunk Connect for Kafka para obtener la documentación detallada del producto. El paquete Splunk Connect for Kafka se instala en un clúster de Kafka Connect del tamaño adecuado (fuera de Splunk), donde puede suscribirse a temas según la configuración y enviar eventos consumidos empleando el HEC para indexarse:

Topología de recopilación de datos empleando Kafka y HEC

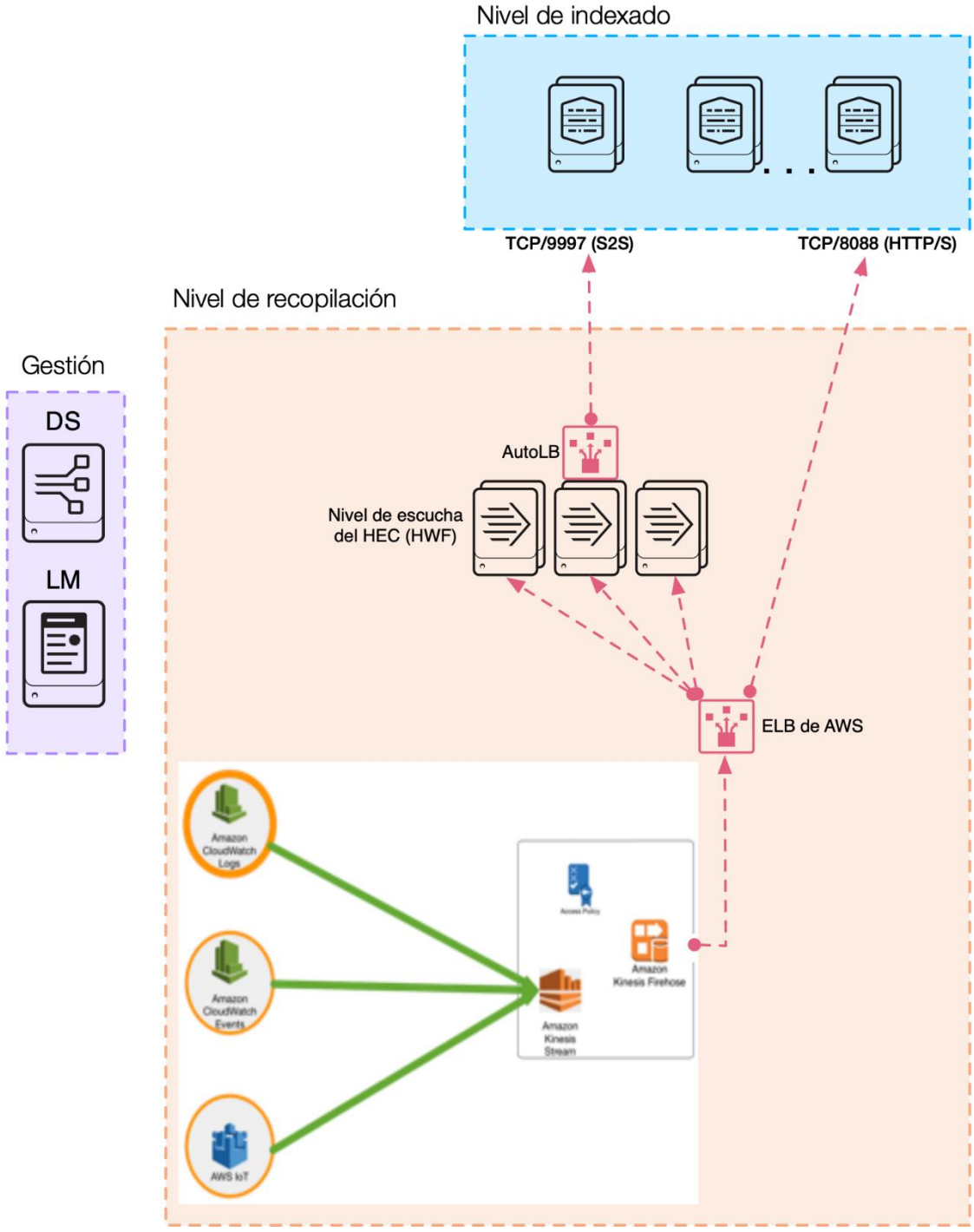


El diagrama muestra Publicadores de Kafka enviando mensajes al bus de Kafka. Las tareas alojadas en el clúster de Kafka Connect consumen esos mensajes a través de Splunk Connect for Kafka y envían los datos al servicio de escucha del HEC empleando un equilibrador de carga de red. De nuevo el servicio de escucha del HEC puede alojarse directamente en los indexadores o en un nivel de escucha del HEC dedicado. Consulte la sección del HEC para ver los detalles. Solo se requieren componentes del nivel de gestión si se implementa un nivel de HF dedicado para alojar escuchas del HEC.

(KINESIS) Consumo de datos de registro desde Amazon Kinesis Firehose

Splunk y Amazon han implementado una integración entre Kinesis y el HEC de Splunk que le permite transmitir datos desde AWS directamente a un extremo del HEC, que es configurable a través de su consola de AWS. Esto se complementa con el complemento de Splunk para Kinesis Firehose que proporciona conocimiento con cumplimiento de CIM para diversas fuentes de datos que se originan en AWS.

Topología de recopilación de datos empleando Amazon Kinesis



El diagrama muestra las fuentes de registros AWS enviándose empleando una transmisión de Kinesis a Firehose que, con la configuración adecuada, envía los datos al servicio de escucha del HEC a través de un ELB de AWS. De nuevo el servicio de escucha del HEC puede alojarse directamente en los indexadores o en un nivel de escucha del HEC dedicado. Consulte la sección del HEC para ver los detalles.

Solo se requieren componentes del nivel de gestión mostrados si se implementa un nivel de HF dedicado para alojar escuchas del HEC.

(MEDICIONES) Recopilación de mediciones

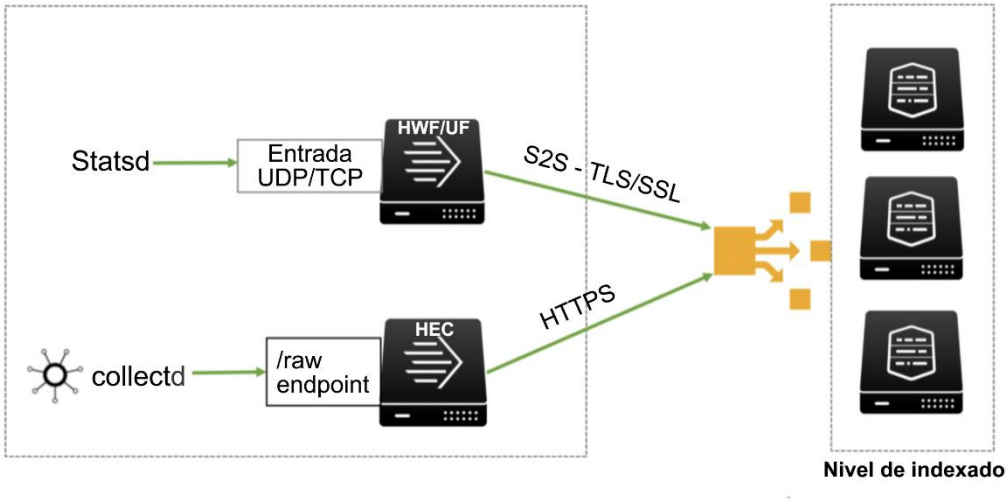
Splunk tiene la capacidad de recibir y recopilar datos de rendimiento del sistema y las aplicaciones, o datos de mediciones, desde una variedad de software de otros fabricantes. Las mediciones en la plataforma Splunk utilizan un tipo de índice personalizado que está optimizado para el almacenamiento y recuperación de mediciones.

Existen varias formas de consumir datos de mediciones y el método de recopilación se basa en la tecnología empleada. La forma más común de las recopilaciones de mediciones es con la forma de un centinela de software, como **collectd**, **statsd** o utilizando un archivo de datos de mediciones personalizado y una configuración válida para la fuente de datos.

Existen principalmente dos métodos para aportar mediciones a Splunk cuando se utilizan agentes como **statsd** y **collectd**. Ya sea utilizando una entrada ***TCP/UDP directa*** o a través del **HEC**.

El uso del **HEC** se considera una práctica recomendada debido a la resilencia y la capacidad de ampliación del extremo del **HE**, y la capacidad de ampliar horizontalmente el nivel de recopilación fácilmente.

Topología de recopilación de datos de mediciones



Statsd admite en estos momentos el transporte **UDP** y **TCP**, que puede utilizar como una entrada directa en un reenviador o indexador de Splunk. No obstante, no es una práctica recomendada enviar tráfico TCP/UDP directamente a los reenviadores en producción, ya que la arquitectura no es resilente y es proclive a la pérdida de eventos (consulte la recopilación Syslog) provocada por los reinicios necesarios del reenviador de Splunk.

Consideraciones sobre la alta disponibilidad (HA) para componentes del nivel de reenvío

Existe un concepto común de alta disponibilidad (HA) en el mundo digital. No obstante, dependiendo de la organización, el significado puede variar y estar más en línea con la recuperación de desastres (DR) en contraposición a la alta disponibilidad. Estos dos conceptos, aunque son parecidos, tienen significados diferentes. HA es una característica de un sistema, que apunta a garantizar un nivel acordado de rendimiento operativo, normalmente tiempo de disponibilidad, para un periodo más alto del normal. DR conlleva una serie de directrices, herramientas y procedimientos para permitir la recuperación o la continuación de la infraestructura tecnológica vital y de los sistemas tras un desastre.

Los siguientes elementos describen las diversas formas de HA en el nivel de agregación/intermedio:

Nivel intermedio

- Para clientes con implementaciones de nivel de agregación/intermedio, la alta disponibilidad de los reenviadores es vital. En la capa de aplicaciones, no dispone en estos momentos de un método de admisión nativa para HA. Hay otras estrategias para proporcionar HA en el nivel del sistema operativo que no son nativas de Splunk. Las soluciones comunes incluyen VMWare VMotion, Grupos de ampliación automática de AWS y Agrupación en clústeres de Linux. Consulte a su Arquitecto de Splunk para tratar otras opciones de diseño disponibles para usted.
- Para entornos con requisitos de alta disponibilidad para un nivel del HEC dedicado, es una práctica recomendada utilizar un equilibrador de carga de tráfico de red (NTLB) como NGINX delante de los reenviadores de alta intensidad múltiples de Splunk. Esto proporciona la ventana de un máximo rendimiento, capacidad de ampliación y disponibilidad. Tiene un conjunto dedicado de instancias del recopilador de eventos HTTP cuyo único trabajo es recibir y reenviar datos. Puede añadir más instancias del HEC si tener necesariamente que añadir más indexadores. Si sus indexadores se convierten en un cuello de botella, agregue indexadores adicionales.
- Para entornos con un requisito de HA para la recopilación de syslog, es una práctica recomendada utilizar múltiples servidores Syslog servidos por una dirección IP (virtual) de clúster alojada por una solución de equilibrio de carga, como HAProxy o F5 para proporcionar el máximo rendimiento, capacidad de ampliación y disponibilidad. Tiene un conjunto dedicado de instancias de Splunk cuyo único trabajo es recibir y reenviar datos. Puede añadir más instancias si tener necesariamente que añadir más indexadores. Si sus indexadores se convierten en un cuello de botella, agregue indexadores adicionales.

Nivel de reenvío

- En el nivel de reenvío (extremo), HA para el agente en sí es dependiente del sistema operativo subyacente. Como mínimo, debería garantizar que cualquier servicio que implemente funciones de reenvío se reinicie automáticamente cuando el sistema operativo anfitrión se reinicie. A parte de esto, las prácticas recomendadas para los reenviadores deberían conllevar la configuración y el uso adecuados de AutoLB desde los reenviadores a múltiples indexadores. Esto también implica el uso del reconocimiento de los indexadores para garantizar que los datos llegan al nivel de indexado.

Paso 3: Aplicar principios de diseño y prácticas recomendadas

A continuación encontrará principios de diseño y prácticas recomendadas separados por nivel de implementación.

Niveles de implementación

Los principios de diseño de las SVA cubren todos los niveles de implementación siguientes:

Nivel	Definición
Búsqueda	• Cabezas de búsqueda
Indexación	• Indexadores
Recopilación	• Reenviadores • Entradas modulares • Red • HEC (Recopilador de eventos HTTP) • etc
Gestión / Utilidad	• CM • DS • LM • DMS • SHC-D

Alineación de su topología con prácticas recomendadas

Tendrá que tener sus requisitos y topología en cuenta para seleccionar los principios de diseño y prácticas recomendadas apropiados para su implementación. Por ello, debería considerar las prácticas recomendadas únicamente después de completar los Pasos 1 y 2 del proceso de selección de Arquitecturas Validadas de Splunk anterior.

Prácticas recomendadas: Recomendaciones específicas de nivel

A continuación encontrará recomendaciones de principios de diseño y prácticas recomendadas para cada nivel de implementación. Cada principio de diseño afianza uno o más pilares de las Arquitecturas Validadas de Splunk. Disponibilidad, Rendimiento, Capacidad de ampliación, Seguridad y Capacidad de gestión

Recomendaciones para el nivel de búsqueda

PRINCIPIOS DE DISEÑO / PRÁCTICAS RECOMENDADAS (Sus requisitos determinarán qué prácticas se le aplican)		PILARES DE LAS SVA				
		DISPONIBILIDAD	RENDIMIENTO	CAPACIDAD DE AMPLIACIÓN	SEGURIDAD	CAPACIDAD DE GESTIÓN
1	Mantener el nivel de búsqueda cerca (en términos de red) del nivel de indexado *Cualquier demora entre el nivel de búsqueda e indexación tendrá una repercusión directa sobre el rendimiento de las búsquedas*		✓			
2	Evitar el uso de múltiples cabezas de búsqueda independientes *Las cabezas de búsqueda independientes no permiten compartir los artefactos de Splunk creados por usuarios. También no tienen buena capacidad de ampliación con respecto a la utilización de los recursos en el nivel de búsqueda. A no ser que haya una necesidad específica de tener entornos de cabezas de búsqueda aisladas, hay una opción mejor para ampliar.*	✓		✓	✓	✓
3	Explotar la agrupación en clústeres de cabezas de	✓		✓		

	búsqueda cuando se amplía el nivel de búsqueda *Un clúster de cabezas de búsqueda replica los artefactos de usuarios en todo el clúster y permite la programación de cargas de trabajo de búsqueda inteligente en todos los miembros del clúster. También proporciona una solución de alta disponibilidad.*					
4	Reenviar todos los registros internos de cabezas de búsqueda al nivel de indexado *Todos los datos indexados deben almacenarse en el nivel de indexado únicamente. Esto elimina la necesidad de proporcionar almacenamiento de alto rendimiento en el nivel de cabezas de búsqueda y simplifica la gestión. Nota: Esto también se aplica a otras funciones de Splunk.*		✓			✓
5	Considerar el uso de la autenticación LDAP siempre que sea posible *La gestión centralizada de las identidades de usuarios para fines de autenticación es una práctica recomendada empresarial general, simplifica la gestión de su implementación de Splunk y aumenta la seguridad.*				✓	✓
6	Garantizar núcleos suficientes para cubrir las necesidades de búsquedas simultáneas *Cada búsqueda requiere la ejecución de un núcleo de CPU. Si no hay núcleos disponibles para ejecutar una búsqueda, esta se pondrá en cola, dando como resultados demoras para el usuario. Nota: Aplicable también al nivel de indexado.*	✓	✓	✓		

7	Utilizar plazos de tiempo de búsqueda programados si es posible / carga de búsquedas programadas suave *A menudo, las búsquedas programadas se ejecutan en puntos específicos del tiempo (a la hora, 5/15/30 minutos después de la hora, a medianoche). Proporcionar un plazo de tiempo en el que se pueda ejecutar su búsqueda ayuda a evitar puntos calientes de simultaneidad.*		✓	✓		
9	Limitar el número de clústeres de cabezas de búsqueda distintos de modo que no se sobrecargue el nivel de indexado *La carga de trabajo de las búsquedas solo puede regirse automáticamente dentro de un entorno SH. Las SHC independientes tienen el potencial de crear más carga de trabajo de búsqueda simultánea de la que puede procesar el nivel de indexado (punto de búsqueda). Lo mismo se aplica para la planificación cuidadosa del número de cabezas de búsqueda autónomas.*	✓		✓		
10	Cuando se construyan agrupaciones en clúster de cabezas de búsqueda, utilizar un número de nodos impar (3, 5, 7, etc,) *La elección del capitán de SHC se realiza utilizando un protocolo basado en mayorías. Un número impar de nodos garantiza que un SHC nunca pueda dividirse en números pares de nodos durante fallos de red.*	✓				✓

Recomendaciones del nivel de indexado

PRINCIPIOS DE DISEÑO / PRÁCTICAS RECOMENDADAS (Sus requisitos determinarán qué prácticas se le aplican)		PILARES				
		DISPONIBILIDAD	RENDIMIENTO	CAPACIDAD DE AMPLIACIÓN	SEGURIDAD	CAPACIDAD DE GESTIÓN
1	Activar canales paralelos en los servidores capaces para *Funciones de paralelización que activan la explotación de los recursos del sistema disponibles que en caso contrario estarían inactivos. Observe que el rendimiento de E/S debe ser adecuado antes de activar funciones de paralelización de introducción.*		✓	✓		
2	Considerar el uso de discos SSD para volúmenes HOT/WARM y resúmenes *Los discos SSD han alcanzado precios económicos y eliminan cualquier posible limitación de E/S que son a menudo la causa de un rendimiento de las búsquedas no satisfactorio.*		✓			
3	Mantener el nivel de indexado cerca (en términos de red) del nivel de búsqueda *La latencia de red menor posible tendrá un efecto positivo en la experiencia de los usuarios al realizar búsquedas.*		✓			
4	Utilizar replicación de índices cuando se necesita alta disponibilidad de datos históricos / informes *La replicación de índices garantiza copias múltiples*	✓				

	de cada evento en el clúster como protección frente a fallos de puntos de búsqueda. Ajuste el número de copias (factor de replicación) para que coincida con sus SLA.					
5	Garantizar una buena higiene de incorporación (por ej. saltos de línea, extracción de marcas de tiempo, TZ y fuente, tipo de fuente y anfitrión están definidos adecuada y explícitamente para cada fuente de datos) y establecer una supervisión continuada mediante la Consola de supervisión *La configuración explícita de fuentes de datos frente a la dependencia sobre las funciones de detección automática de Splunk se ha demostrado tener un beneficio significativo sobre la capacidad de introducción de datos y la latencia de indexado, especialmente en implementaciones de alto volumen.*		✓	✓		✓
6	Considerar la configuración del parámetro de paralelización de búsquedas en modo por lotes en los indexadores con potencia adicional de procesamiento *Explotar las funciones de paralelización de búsquedas puede tener una repercusión significativa sobre el rendimiento de las búsquedas para ciertos tipos de ellas, y le permite utilizar recursos del sistema que en caso contrario quedarían sin utilizar.*		✓	✓		
7	Supervisar la distribución de datos equilibrada en		✓	✓		✓

	los nodos de indexadores (=puntos de búsqueda). *La distribución uniforme de los datos/eventos en los puntos de búsqueda es un factor crítico que contribuye al rendimiento de las búsquedas y la aplicación de las directrices de retención de datos adecuadas.*					
8	Desactivar la interfaz de usuario web en indexadores en implementaciones distribuidas/agrupadas en clústeres. *No hay necesidad razonable de acceder a la interfaz de usuario web directamente en los indexadores.*		✓		✓	✓
9	Considerar los complementos de tecnología preconstruidos de Splunk para fuentes de datos bien conocidas *En vez de construir su propia configuración para garantizar la higiene de incorporación de los datos para fuentes de datos bien conocidas, los TA proporcionados por Splunk proporcionan un tiempo de generación de valor más rápido y garantizan una implementación óptima.*		✓			✓
10	Supervisar mediciones de indexadores críticas *Splunk proporciona una consola de supervisión que proporciona mediciones de rendimiento clave sobre el rendimiento de su nivel de indexado. Esto incluye la utilización de la CPU y la memoria, así como mediciones detalladas de componentes internos de Splunk (procesos, canales, colas, búsqueda).*	✓	✓			

Recomendaciones del nivel de recopilación

	PRINCIPIOS DE DISEÑO / PRÁCTICAS RECOMENDADAS (Sus requisitos determinarán qué prácticas se le aplican)	PILARES				
		DISPONIBILIDAD	RENDIMIENTO	CAPACIDAD DE AMPLIACIÓN	SEGURIDAD	CAPACIDAD DE GESTIÓN
1	Utilizar UF para reenviar datos siempre que sea posible. El uso del reenviador de alta intensidad debería limitarse a casos de uso que lo requieran. *autoLB incorporado, capacidad de reinicio, configurable de manera centralizada, pequeña demanda de recursos*		✓			✓
2	Utilizar al menos 2x canales de reenvío intermedios a indexadores al canalizar muchos UF *El multiplexado de un gran número de reenviadores de extremos por un número reducido de reenviadores intermedios afecta a la distribución uniforme de los eventos entre indexadores, lo que afecta al rendimiento de las búsquedas.* ***Solo implemente reenviadores intermedios si es absolutamente necesario.***	✓	✓			
3	Considerar asegurar el tráfico UF-IDX empleando SSL				✓	
4	Utilizar LB de Splunk nativo para diseminar los datos en el nivel de indexado	✓		✓		

	Los equilibradores de carga de red no se admiten en estos momentos entre reenviadores e indexadores.					
5	Utilizar servidores syslog dedicados para la recopilación syslog *Los servidores syslog pueden hacer persistir el tráfico TCP/UDP en disco en base a la fuente y permiten la adecuada configuración del tipo de fuente para su procesamiento con un reenviador universal. Los reinicios del reenviador requerido no provocarán pérdidas de datos.*					
6	Utilizar el HEC para la recopilación sin agentes (en vez de TCP/UDP nativo) *El Recopilador de eventos HTTP (HEC) es un servicio de escucha que permite la publicación de eventos a través del protocolo HTTP[/S]. Puede activarse directamente en los indexadores, o configurarse en un nivel de reenviador de alta intensidad, ambos servidos por un equilibrador de carga.*					

Recomendaciones del nivel de gestión / utilidad

	PRINCIPIOS DE DISEÑO / PRÁCTICAS RECOMENDADAS (Sus requisitos determinarán qué prácticas se le aplican)	PILARES				
		DISPONIBILIDAD	RENDIMIENTO	CAPACIDAD DE AMPLIACIÓN	SEGURIDAD	CAPACIDAD DE GESTIÓN
1	Considerar la consolidación de LM, CM, SHC-D y MC en una única instancia para entornos pequeños *Estas funciones de servidor tienen demandas de recursos muy reducidas y son buenas candidatas para su colocación. En clústeres de indexadores de mayor tamaño, el CM puede requerir un servidor dedicado para gestionar eficientemente el clúster.*					✓
2	Considerar una instancia separada para DS para implementaciones de medio a gran tamaño *Una vez haya un número significativo de reenviadores gestionados a través del Servidor de implementación, las necesidades de recursos aumentarán hasta que se requiera un servidor dedicado para mantener el servicio.*					✓
3	Considere múltiples DS detrás de LB para implementaciones de tamaño masivo *Nota: Esto puede requerir ayuda de los Servicios Profesionales de Splunk para la instalación y configuración adecuadas.*	✓		✓		✓
4	Determinar si phoneHomeIntervalInSecs de DS puede respaldarse dentro del valor			✓		

	predeterminado de 60 segundos *Un intervalo telefónico más prolongado tendrá un efecto positivo en la capacidad de ampliación del DS.*					
5	Utilizar un DS dedicado/asegurado para evitar la explotación de clientes a través de la implementación de aplicaciones *Cualquiera con acceso al Servidor de implementación puede modificar la configuración de Splunk gestionada por ese DS, incluyendo la potencial implementación de aplicaciones malintencionadas en extremos de reenviadores. Asegurar esta función apropiadamente es prudente.*				✓	
6	Utilizar la Consola de supervisión (MC) para supervisar el estado de su implementación y alertas sobre problemas de estado *La Consola de supervisión proporciona un conjunto específico de soluciones de supervisión preconstruidas de Splunk y contiene alertas de plataforma ampliables que pueden notificarle sobre una degradación del estado de su entorno.*	✓	✓			✓

Resumen y siguientes pasos

Este documento técnico ha proporcionado una introducción general a las Arquitecturas Validadas de Splunk Una Arquitectura Validada garantiza que los requisitos de su organización se están cumpliendo de la manera más rentable, gestionable y ampliable posible. Las SVA ofrecen prácticas recomendadas y principios de diseño construidos sobre los siguientes pilares fundamentales:

- Disponibilidad
- Rendimiento
- Capacidad de ampliación
- Seguridad
- Capacidad de gestión

Este documento técnico también ha tratado el proceso de selección de 3 pasos de las Arquitecturas Validadas de Splunk: 1) Definición de los requisitos, 2) Selección de una topología 3) Aplicación de los principios de diseño y las prácticas recomendadas. Ahora que está familiarizado con los múltiples beneficios de las Arquitecturas Validadas de Splunk, esperamos que esté listo para avanzar en el proceso de elección de una topología de implementación adecuada para su organización.

Siguientes pasos

¿Qué viene después de elegir una Arquitectura Validada? Los siguientes pasos de su trayectoria a un entorno de funcionamiento incluyen:

Personalizaciones

- Tenga en cuenta cualquier personalización necesaria que su topología podría necesitar para cumplir requisitos específicos.

Modelo de implementación

- Decida el modelo de implementación (desde cero, virtual, nube)

Sistema

- Seleccione su tecnología (servidores, almacenamiento, sistemas operativos) según los requisitos del sistema de Splunk.

Cálculo del tamaño

- Recopile todos los datos relevantes que necesitará para calcular el tamaño de su implementación (introducción de datos, volumen de búsquedas esperado, necesidades de retención de datos, replicación, etc.) Cálculo de tamaño de almacenamiento de Splunk (https://splunk-sizing.appspot.com/) es una de las herramientas disponibles.

Plantilla

- Evalúe sus necesidades de plantilla para implementar y gestionar su implementación. Esto es una parte esencial para la construcción de un Centro de distinción de Splunk.

Estamos aquí para ayudarle en todo el proceso de las Arquitecturas Validadas y con los siguientes pasos. Tiene la libertad de participar con su Equipo de cuenta de Splunk con cualquier pregunta que pueda tener. Su Equipo de cuenta tendrá acceso al conjunto completo de recursos técnicos y arquitectónicos dentro de Splunk y estará encantado de proporcionarle más información.

¡Feliz Splunking!

Apéndice

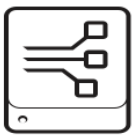
Esta sección contiene información de referencia adicional empleada en las SVA.

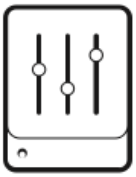
Apéndice "A": Pilares de las SVA explicados

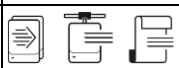
Pilar	Descripción	Objetivos principales / Principios de diseño
Disponibilidad	La capacidad de funcionamiento continuo y de recuperarse de cortes o interrupciones planificados y no planificados.	1. Eliminar puntos de fallo únicos / Agregar redundancia 2. Detectar cortes/interrupciones planificados y no planificados 3. Tolerar interrupciones planificadas y no planificadas, idealmente de forma automática 4. Planificar actualizaciones graduales
Rendimiento	La capacidad de utilizar de forma efectiva los recursos disponibles para mantener le nivel óptimo del servicio bajo patrones de uso cambiantes.	1. Agregue hardware para mejorar el rendimiento: computación, almacenamiento, memoria 2. Elimine los cuellos de botella de abajo a arriba 3. Explote todos los medios de procesamiento simultáneo 4. Explote la localidad (por ej. la reducción al mínimo de la distribución de componentes) 5. Optimice para el caso común (regla 80/20) 6. Evite la generalización innecesaria 7. Computación del cambio de tiempo (precomputación, computación lenta, computación para compartir/lotes) 8. Intercambiar la certeza y la precisión por el tiempo (aleatorización, muestreo)
Capacidad de ampliación	La capacidad de garantizar que el sistema está diseñado para ampliarse en todos los niveles y controlar las cargas de trabajo aumentadas de manera efectiva.	1. Ampliación vertical y horizontal 2. Separación de los componentes funcionales que necesitan ampliarse individualmente

Pilar	Descripción	Objetivos principales / Principios de diseño
		3. Minimización de las dependencias entre componentes 4. Diseño de crecimiento futuro conocido lo antes posible 5. Introducción de una jerarquía en el diseño general del sistema
Seguridad	La capacidad de garantizar que el sistema está diseñado para proteger los datos así como las configuraciones/activos mientras sigue entregando valor.	1. Diseño de un sistema seguro desde el principio 2. Empleo de protocolos vanguardistas para todas las comunicaciones 3. Permitir el acceso amplio y granular a datos de eventos 4. Empleo de una autenticación centralizada 5. Implementación de procedimientos de auditoría 6. Reducción de los ataques o el uso malintencionado
Capacidad de gestión	La capacidad de garantizar que el sistema es operativo y gestionable de forma centralizada en todos sus niveles.	1. Proporcionar una función de gestión centralizada 2. Gestión del ciclo de vida de los objetos de configuración (control de fuente) 3. Medición y supervisión/perfil del uso de la aplicación (Splunk) 4. Medición y supervisión del estado del sistema

Apéndice "B": Componentes de topología

Nivel	Componente	Icono	Descripción	Notas
Gestión	Servidor de implementación (DS)		La implementación del servidor gestiona la configuración del reenviador.	Debe implementarse en una instancia dedicada. Puede virtualizarse para una fácil recuperación de los fallos.

Nivel	Componente	Icono	Descripción	Notas
	Maestro de licencias (LM)		Ortos componentes de Splunk requieren el Maestro de licencias para activar las funciones con licencia y realizar un seguimiento del volumen de introducción de datos diario.	La función de maestro de licencias tiene requisitos de capacidad y disponibilidad mínimos y puede colocarse con otras funciones de gestión. Puede virtualizarse para una fácil recuperación de los fallos.
	Consola de supervisión (MC)		La Consola de supervisión proporciona paneles indicadores para la supervisión del uso y el estado de su entorno. También contiene un número de alertas de plataforma preempaquetadas que pueden personalizarse para proporcionar notificaciones de problemas operativos.	En entornos agrupados en clústeres, la MC puede colocarse con el Nodo maestro, además del Maestro de licencias y la función de Servidor de implementación en implementaciones no agrupadas en clústeres. Puede virtualizarse para una fácil recuperación de los fallos.
	Maestro de clúster (CM)		El Maestro de clúster es el coordinador requerido para toda la actividad en una implementación agrupada en clústeres.	En clústeres con un gran número de depósitos de índices (alto volumen/retención de datos), el Maestro de clúster probablemente requerirá un servidor dedicado en el que ejecutarse. Puede virtualizarse para una fácil recuperación de los fallos.
	Implementador de clúster de cabezas de búsqueda (SHC-D)		El Implementador de clúster de cabezas de búsqueda se necesita para el arranque de un SHC y gestionar la configuración de Splunk implementada en el clúster.	El SHC-D no es un componente en tiempo de ejecución y tiene requisitos del sistema mínimos. Puede colocarse con otras funciones de gestión. Nota: Cada SHC requiere su propia función implementadora de SHC. Puede virtualizarse para una fácil recuperación de los fallos.

Nivel	Componente	Icono	Descripción	Notas
Búsqueda	Cabeza de búsqueda (SH)		La cabeza de búsqueda proporciona la interfaz de usuario para los usuarios de Splunk y coordina la actividad de búsquedas programadas.	Las cabezas de búsqueda son instancias de Splunk dedicadas en implementaciones distribuidas. Las cabezas de búsqueda pueden virtualizarse para una fácil recuperación de fallos, siempre que se implementen con los recursos de CPU y memoria adecuados.
	Clúster de cabezas de búsqueda (SHC)		Un clúster de cabezas de búsqueda es un conjunto de al menos tres cabezas de búsqueda agrupadas en un clúster. Proporciona capacidad de ampliación horizontal para el nivel de cabezas de búsqueda y una recuperación de fallos de usuario transparente en caso de interrupciones.	Los clústeres de cabezas de búsqueda son servidores dedicados con especificaciones del sistema idénticas de manera ideal. Los miembros del clúster de cabezas de búsqueda pueden virtualizarse para una fácil recuperación de fallos, siempre que se implementen con los recursos de CPU y memoria adecuados.
Indexación	Indexador		Los indexadores son el corazón y el alma de Splunk. Procesa e indexan los datos entrantes y también sirven como puntos de búsqueda para cumplir las solicitudes de búsqueda iniciadas en el nivel de búsqueda.	Los indexadores deben estar siempre en servidores dedicados en implementaciones distribuidas o agrupadas en clústeres. En una implementación de un único servidor, el indexador también proporciona la interfaz de usuario de búsqueda y las funciones maestras de licencia. Los indexadores tienen su mejor rendimiento en servidores creados desde cero o en máquinas virtuales dedicadas de alto rendimiento, si se pueden garantizar los recursos adecuados.
Recopilación de datos	Los reenviadores y otros componentes de recopilación de datos		El icono general para cualquier componente involucrado en la recopilación de datos.	Esto incluye los reenviadores universales y de alta intensidad, las entradas de datos de red y otras formas de recopilación de datos (HEC, Kafka, etc.)