

Google Search Appliance

Arquitecturas de implementación

Agosto de 2014



© 2014 Google

Arquitecturas de implementación

Google Search Appliance (GSA) ofrece varias funciones que permiten una implementación en distintas arquitecturas para satisfacer diversos requisitos de alta disponibilidad, rendimiento y escala de contenido. En este documento se proporcionan directrices para usar funciones de arquitectura, así como otros componentes de la implementación compatibles, para diseñar una arquitectura de soluciones GSA.

Acerca de este documento

La información y recomendaciones que se exponen en este documento se recopilamos gracias al trabajo que realizamos con distintos clientes y entornos del sector. Nos gustaría agradecer a nuestros clientes y partners que hayan compartido sus experiencias y puntos de vista con nosotros.

Temas tratados	En este documento se describen algunos ejemplos típicos de arquitecturas de implementación y cómo se puede implantar el dispositivo GSA para hacer frente a las necesidades de una comunidad de usuarios de gran envergadura y producción, y donde los fallos en el sistema no son una opción
Funciones y responsabilidades	<ul style="list-style-type: none">• Administradores de GSA: implementar y configurar los dispositivos GSA para mostrar contenido de manera óptima a la comunidad de usuarios que usan el sistema de búsqueda de la empresa• Administradores de red: confirmar la capacidad y disponibilidad de la red para gestionar el tráfico que puedan imponer las distintas configuraciones de la red corporativa• Propietarios del contenido: acceso de intermediario a las distintas fuentes de contenido a las que GSA deberá acceder para facilitar el rastreo, la indexación y la búsqueda
Entorno de TI	GSA, SNMP, distribuidor de carga, interruptor de red, servidor proxy de búsqueda
Fases de la implementación	Planificación de la implementación
Otros recursos	<ul style="list-style-type: none">• Learngsa.com proporciona recursos formativos sobre GSA• En la documentación de producto de GSA se ofrece toda la información sobre la solución• El Portal de asistencia de Google for Work proporciona acceso a la asistencia de Google

Índice

[Acerca de este documento](#)

[Capítulo 1 Componentes de la arquitectura](#)

[Funciones básicas de arquitectura del dispositivo GSA](#)

[Duplicación de GSA](#)

[Agrupación de GSA](#)

[rastreo y presentación de resultados distribuidos de GSA](#)

[Combinar la duplicación, la agrupación y el rastreo y presentación de resultados distribuidos de GSA](#)

[Capítulo 2 Diseñar una arquitectura para lograr una alta disponibilidad](#)

[Alta disponibilidad de GSA](#)

[Supervisar y detectar fallos](#)

[Alta disponibilidad de los conectores](#)

[Alta disponibilidad de los conectores para la transferencia de contenido](#)

[Alta disponibilidad de los mecanismos de seguridad](#)

[Capítulo 3 Diseñar una arquitectura para lograr un alto rendimiento](#)

[Distribuir la carga de GSA para optimizar el rendimiento de las consultas](#)

[Rendimiento de los mecanismos de seguridad](#)

[Consideraciones sobre el rendimiento de las funciones de GSA](#)

[Capítulo 4 Diseñar una arquitectura para índices de gran escala](#)

[Estrategias de rastreo de GSA](#)

Capítulo 1 Componentes de la arquitectura

Para diseñar la arquitectura de una implementación de Google Search Appliance hay que tener en cuenta varias funciones de GSA, así como los puntos de integración con los componentes de una implementación de GSA. En este capítulo se ofrece una descripción general de la función que cada uno de los componentes, como los conectores y los mecanismos de seguridad de GSA, tiene en la arquitectura de una implementación de este dispositivo.

Funciones básicas de arquitectura del dispositivo GSA

GSA proporciona tres funciones principales que se pueden implementar para hacer frente a los requisitos de alta disponibilidad, rendimiento y escalabilidad:

- [Duplicación de GSA](#)
- [Agrupación de GSA](#)
- [Rastreo y presentación de resultados distribuidos de GSA](#)

Duplicación de GSA

Proporciona la réplica de un índice de búsqueda y de la mayoría de los ajustes de configuración en uno o varios dispositivos de búsqueda réplica.

Las arquitecturas de duplicación se pueden diseñar tanto para configuraciones activo-pasivas como activo-activas y se pueden usar para proporcionar una alta disponibilidad y alto rendimiento en la implementación de GSA.

Si desea obtener más información sobre la duplicación de GSA, consulte estos artículos del Centro de ayuda de GSA: [Configurar la duplicación de GSA](#) y [Especificaciones y límites: duplicación de GSA](#).

Agrupación de GSA

Agrupación de GSA permite configurar un grupo de dispositivos de búsqueda GSA para que los documentos indexados en los diferentes dispositivos puedan aparecer como resultado de una misma consulta de búsqueda. Cada uno de los dispositivos de búsqueda de la configuración indexa distintos grupos de documentos y están configurados con sus propias colecciones, interfaces y otras funciones que puede ajustar un administrador. Cuando un usuario realiza una búsqueda, los dispositivos de búsqueda se comunican para combinar los resultados de los distintos índices.

El entorno GSA agrupado se suele emplear cuando hay que proporcionar servicios de búsqueda e indexación para un corpus de documentos considerable y que un solo dispositivo Google Search Appliance no puede alojar o bien para agrupar corpus repartidos en distintos puntos de la geografía.

Uno de los dispositivos de búsqueda de la configuración se designa como dispositivo de búsqueda o nodo principal, mientras que los otros serán dispositivos o nodos secundarios.

Si desea obtener más información sobre la duplicación de GSA, consulte estos artículos del Centro de ayuda de GSA: [Configurar la duplicación de GSA](#) y [Especificaciones y límites: duplicación de GSA](#).

Rastreo y presentación de resultados distribuidos de GSA

El rastreo y presentación de resultados distribuidos (DC+S, Distributed Crawling and Serving) es una función de Google Search Appliance que permite ampliar la capacidad para documentos del dispositivo de búsqueda. El rastreo y presentación de resultados distribuidos permite que varios dispositivos de búsqueda funcionen como un único dispositivo. Después de habilitar DC+S, el rastreo, la indexación y la presentación de resultados se configuran en un dispositivo de búsqueda principal de administración denominado "admin master". Los demás dispositivos son secundarios o "non-master".

Por ejemplo, si hay dos dispositivos de búsqueda con licencia para rastrear 100 millones de documentos cada uno, al habilitar DC+S, los dispositivos podrán rastrear un total de 200 millones de documentos.

Si desea obtener más información sobre la función DC+S de GSA, consulte estos artículos del Centro de ayuda de GSA: [Configurar el rastreo y presentación de resultados distribuidos](#) y [Especificaciones y límites: rastreo y presentación de resultados distribuidos de GSA](#).

Combinar la duplicación, la agrupación y el rastreo y presentación de resultados distribuidos de GSA

A partir de la versión 7.2 de GSA, se admiten las siguientes combinaciones de arquitectura:

La duplicación de GSA se puede usar en:

- **Red de duplicación GSA** independiente con un dispositivo principal vinculado a uno o varios dispositivos réplica
- **Red de agrupación de GSA**
- **Red de rastreo y publicación de resultados distribuido de GSA**, donde se pueden implantar dispositivos réplica para cada uno de los dispositivos secundarios

La agrupación de GSA se puede usar en:

- **Red de agrupación de GSA** independiente con un dispositivo principal vinculado a uno o varios nodos
- **Red de agrupación de GSA con duplicación de GSA** de cualquier nodo

El rastreo y presentación de resultados distribuidos de GSA se puede usar en:

- **Red de rastreo y presentación de resultados distribuidos de GSA** con un dispositivo principal vinculado a uno o varios dispositivos secundarios
- **Red de rastreo y presentación de resultados distribuidos de GSA con duplicación de GSA** para todos los nodos

Capítulo 2 Diseñar una arquitectura para lograr una alta disponibilidad

Alta disponibilidad de GSA

Para que GSA pueda ofrecer una alta disponibilidad, deberá implementar un distribuidor de carga delante de varios dispositivos de búsqueda que incluyan el mismo índice de búsqueda. A través de la supervisión y de la detección de fallos, un evento del sistema de conmutación por error implicaría modificar el distribuidor de carga para que redirija el tráfico a un dispositivo GSA réplica. Esto puede hacerse de forma manual o automática (según la solución de distribución de carga o de supervisión que se utilice). GSA no ofrece funciones automáticas de distribución de carga ni de conmutación por error. Por lo tanto, estos eventos deben gestionarse mediante componentes externos.

Para lograr un índice de búsqueda uniforme en varios dispositivos, se suele usar la duplicación de GSA. Sin embargo, si esto no es posible (p. ej., porque la conectividad de red entre los dispositivos es insuficiente), habrá que configurar cada dispositivo de búsqueda para que rastree, indexe y muestre el mismo contenido. Así se garantiza una experiencia de usuario coherente en caso de conmutación por error.

Nota sobre las licencias: Si el tráfico solo se dirige al dispositivo GSA principal, la arquitectura se conoce como configuración *activo-pasiva*. En este caso, la licencia de tipo "producción" de GSA solo se necesita para el dispositivo principal, mientras que los dispositivos réplica pueden usar una licencia de tipo "secundaria".

Si desea obtener más información sobre la distribución de carga de GSA, consulte este artículo del Centro de ayuda de GSA: [Configurar dispositivos de búsqueda para la distribución de carga o conmutación por error](#).

Supervisar y detectar fallos

Supervisar GSA mediante SNMP

Google Search Appliance permite la integración de SNMP (Simple Network Management Protocol o protocolo de gestión de redes simple) para que pueda recibir mensajes cuando el estado operativo de Google Search Appliance cambia. El dispositivo de búsqueda escucha solicitudes SNMP en el puerto 161 UDP; es compatible con las versiones 1, 2 y 3 de SNMP.

El servidor SNMP de Google Search Appliance proporciona un subconjunto de la información del estado acerca del dispositivo de búsqueda que está disponible en la Consola del administrador. El dispositivo de búsqueda admite los comandos `Get` y `GetNext` de SNMP. En cambio, no admite `Trap` ni definir los valores a través de `Set` de SNMP.

Nota: Esta función se suele configurar únicamente en entornos donde SNMP ya se usa para gestionar otros dispositivos de la red, como routers, interruptores, servidores de aplicaciones o servidores de

almacenamiento. En caso de que no se utilice SNMP, piense en la posibilidad de implementar un sistema de supervisión de servidores personalizado (véase más abajo).

Si desea obtener información más detallada al respecto, consulte la página de ayuda de la Consola del administrador de GSA: **Administración** > [Configuración SNMP](#).

Usar un sistema de supervisión personalizado

Si no se dispone de un sistema de supervisión para comprobar el estado del dispositivo GSA, se puede configurar un sistema personalizado mediante una aplicación web simple. La aplicación web puede usar la API de administración y acceder a la información de estado de GSA. Luego la aplicación web muestra el estado en una página de estado para que el administrador pueda supervisarlos.

La API de administración también se puede usar para supervisar Google Search Appliance y almacenar la información en una base de datos o archivos de registro; esta información se puede consultar y analizar en otro momento.

Si desea consultar las prácticas recomendadas para implementar un sistema de supervisión para las consultas de búsqueda, lea [Supervisar la publicación de resultados de GSA](#).

Alta disponibilidad de los conectores

Los conectores de GSA realizan dos funciones principales que pueden influir en la disponibilidad: 1) descubrimiento y transferencia de contenido; y 2) servicios de seguridad para autenticaciones y autorizaciones.

Alta disponibilidad de los conectores para la transferencia de contenido

Si bien la transferencia de contenido se puede considerar un servicio menos importante en términos de alta disponibilidad, a veces es un requisito en entornos donde es clave que el contenido esté actualizado en todo momento. Si se utiliza un conector, la solución para aplicar una alta disponibilidad al rastreo dependerá del tipo de conector y de la versión de la estructura para conectores (es decir, versiones 3.x o 4.x).

Para la mayoría de los conectores de la estructura para gestor de conectores 3.x (salvo el conector para sistemas de archivos), las interrupciones en el rastreo mediante conector (relacionadas con el propio conector) necesitarán una intervención manual para reiniciar el conector o habilitar otro conector para el rastreo. Dado que estos conectores tienen estado, en la mayoría de las acciones de recuperación cuando se produce un fallo de rastreo del conector deberá reiniciarse (o restablecerse) la transferencia mediante conector. En algunos casos, se puede hacer una copia de seguridad de la información de estado de transferencia del conector (como el archivo de estado XML del conector para SharePoint) para retomar la transferencia desde el estado actual, en lugar de hacer una transferencia completa.

En el caso de los conectores de la estructura para conectores 4.x (y también el conector para sistemas de archivos 3.x), puede conseguir un rastreo de alta disponibilidad implementando instancias de copia de seguridad del conector y colocándolas detrás de un distribuidor de carga. Estos conectores secundarios deberían tener inhabilitadas las funciones de transferencia completa y de descubrimiento. Por lo tanto, solo pueden recuperar ciertos documentos que solicite GSA. De esta forma, si hay un conector inactivo,

el distribuidor de carga podrá dirigir las solicitudes de rastreo a otro conector manteniendo la disponibilidad del rastreo. En este caso, la intervención manual también es necesaria para registrar y descubrir contenido nuevo del conector principal que se ha configurado para la transferencia.

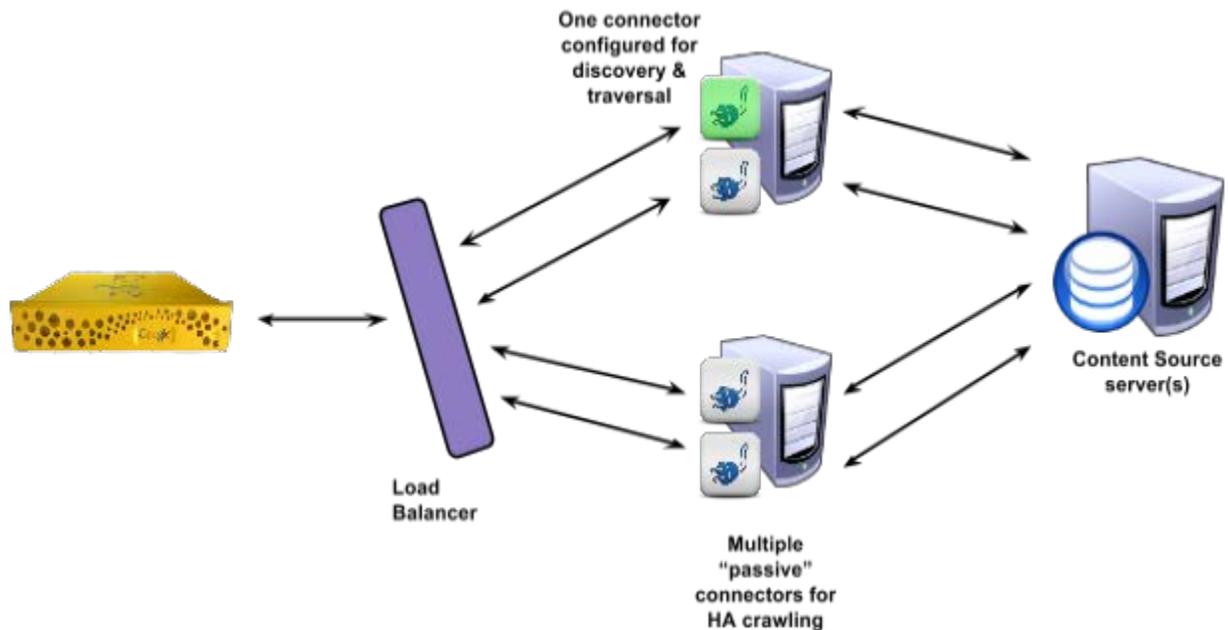


Diagrama 1. Arquitectura de ejemplo para lograr una alta disponibilidad de la transferencia de contenido a través de conectores

Alta disponibilidad de los conectores para los mecanismos de seguridad

Si desea conseguir una alta disponibilidad de los conectores que ofrecen servicios de autenticación y de autorización, implemente varios conectores y utilice un distribuidor de carga para dirigir el tráfico entre ellos. Si desea obtener información más detallada sobre este caso, consulte [Notas de GSA: Introducción a la integración de contenido](#).

Alta disponibilidad de los mecanismos de seguridad

En entornos de búsqueda segura, la alta disponibilidad en los mecanismos de seguridad es un factor clave, ya que las interrupciones en los mecanismos de autenticación y de autorización implican la interrupción de la búsqueda segura. La alta disponibilidad en estos casos se suele aplicar mediante la implementación de un distribuidor de carga delante de varias instancias del mecanismo de seguridad. Por ejemplo, se puede implementar un distribuidor de carga delante de varios conectores para solicitudes de autenticación y autorización o se puede configurar una URL con carga distribuida para sistemas de inicio de sesión único.

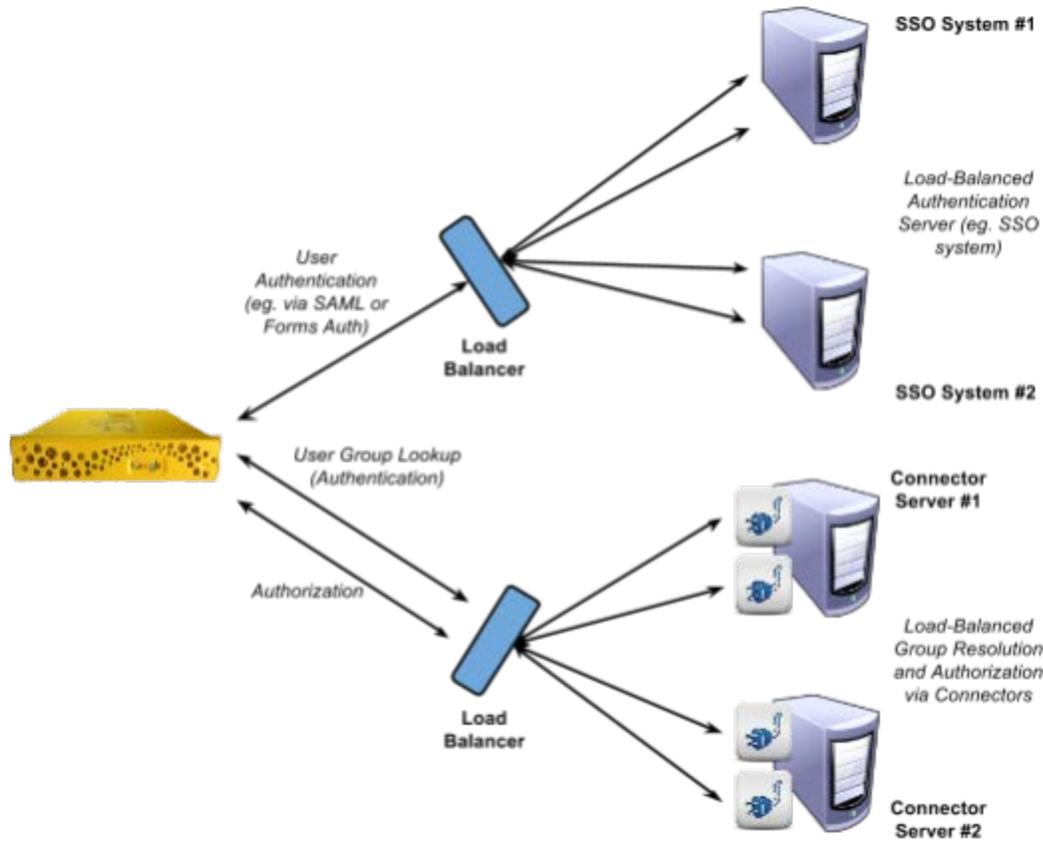


Diagrama 2. Arquitectura de ejemplo para lograr alta disponibilidad y rendimiento en los mecanismos de seguridad

Capítulo 3 Diseñar una arquitectura para lograr un alto rendimiento

Distribuir la carga de GSA para optimizar el rendimiento de las consultas

El número de consultas por segundo (QPS o Queries per Second) que admite un dispositivo de búsqueda depende de cuánto tarda una consulta, el número de solicitudes simultáneas y el tipo de funciones de interfaz que estén habilitadas. Un solo dispositivo admite hasta 50 solicitudes simultáneas, pero generalmente el QPS de un dispositivo es menor, sobre todo en las búsquedas seguras. Si desea obtener más información sobre las solicitudes simultáneas, consulte estos artículos del Centro de ayuda: [Diseñar una solución de búsqueda](#) y [¿Cuántos usuarios simultáneos admite un dispositivo GSA?](#)

Para alcanzar el nivel de rendimiento de consultas deseado, se pueden implementar varios dispositivos detrás del distribuidor de carga, el cual reparte el tráfico de búsqueda en los dispositivos GSA de manera uniforme. Cada aplicación debería incluir el mismo índice, ya sea a través de la duplicación de GSA o mediante el rastreo independiente del mismo contenido. En este caso, los dispositivos GSA se consideran como una configuración *activo-activa* (ya que todos los dispositivos GSA muestran resultados de manera activa); todos los dispositivos deberían tener la licencia de tipo "producción".

En configuraciones de carga distribuida que incluyen búsquedas seguras deberían usarse sesiones cohesivas, ya que todas las sesiones de búsqueda segura están vinculadas a un determinado dispositivo. Por lo tanto, si se dirige un usuario a otro dispositivo de búsqueda después de haberse autenticado, quizá deba hacerlo de nuevo.

Rendimiento de los mecanismos de seguridad

Otro cuello de botella que suele producirse habitualmente y que afecta al rendimiento es al autorizar los resultados de búsqueda, en concreto cuando se utilizan enlaces en tiempo de ejecución (mecanismo en tiempo real que hace las comprobaciones en un sistema de contenido) para determinar los permisos de un documento.

El método de autorización recomendado son los enlaces en tiempo de compilación porque el rendimiento es superior con las comprobaciones de seguridad en GSA. Sin embargo, hay casos en los que es necesario usar los enlaces en tiempo de ejecución porque no es posible extraer las Listas de control de acceso (LCA) o porque los permisos de seguridad de los documentos cambian con frecuencia y se requieren comprobaciones en tiempo real.

Para mejorar el rendimiento de las autorizaciones, se pueden aplicar varias estrategias:

- Si las cargas elevadas degradan el rendimiento de las autorizaciones, implemente varias instancias del mecanismo de autorización con carga distribuida (p. ej., varios proveedores de SAML).
- Si se utiliza un conector para realizar las autorizaciones y también para rastrear y transferir contenido, implemente otros conectores que gestionen las autorizaciones y utilice un servidor proxy o un distribuidor de carga para repartir el tráfico entre las solicitudes de rastreo y de autorización que emita GSA.

Si bien el rendimiento del proceso de autenticación preocupa menos (ya que los usuarios solo lo realizan una vez por sesión), se pueden aplicar las mismas estrategias para optimizar el rendimiento

de los mecanismos de autenticación. En el diagrama 2 puede ver un ejemplo de cómo implementar mecanismos de seguridad de alto rendimiento.

Consideraciones sobre el rendimiento de las funciones de GSA

Si bien GSA proporciona varias funciones que mejoran la experiencia del usuario, cuando diseñe una implementación de GSA de alto rendimiento deberá tener en cuenta cómo afectan dichas funciones al rendimiento:

- Cuando **Navegación dinámica** se usa con búsquedas seguras, GSA debe realizar muchas más solicitudes de autorización. Si esta función está inhabilitada, GSA realiza solicitudes de autorización hasta que obtiene 10 resultados permitidos para el usuario cuando se utilizan enlaces en tiempo de ejecución (suponiendo que se solicitan 10 resultados), y 1.000 si se usan enlaces en tiempo de compilación. Con Navegación dinámica, GSA realizará solicitudes de autorización hasta que obtenga *10.000 resultados permitidos*. Esto afectará al rendimiento de manera considerable si se utilizan enlaces en tiempo de ejecución. También tendrá efectos en los casos en que se utilicen enlaces en tiempo de compilación y los usuarios tengan acceso limitado al contenido. Si desea obtener más información, consulte este artículo del Centro de ayuda: [Los resultados de búsqueda se ralentizan si se habilita Navegación dinámica](#).
- **Agrupación de resultados dinámicos** realiza muchas solicitudes de autorización para determinar las agrupaciones de resultados y también envía una solicitud HTTP adicional a GSA.
- Las **solicitudes de OneBox** se realizan de forma sincronizada y, por tanto, se suman al tiempo de respuesta general, concretamente si las solicitudes se realizan a un sistema de terceros.
- **Sugerencias de consulta** envía solicitudes HTTP adicionales a GSA para determinar sugerencias a medida que el usuario escribe. En el caso de las implementaciones con muchos usuarios, esto puede suponer una carga adicional importante para GSA. Este efecto puede minimizarse ajustando el tiempo de espera tras el cual un evento de tecla activa la solicitud al servicio de Sugerencias de consulta en el dispositivo GSA. Consulte este artículo del Centro de ayuda: [Impacto de las sugerencias de consulta en el rendimiento general de la búsqueda](#).
- Los **filtros de resultados de búsqueda** (p. ej., filtros para carpetas duplicadas, títulos o fragmentos) pueden afectar al rendimiento en entornos de búsqueda segura, porque para mostrar un conjunto de resultados se requiere un mayor número de solicitudes de autorización. Por ejemplo, si un usuario solicita 100 resultados en un entorno de enlaces en tiempo de ejecución y los 100 primeros resultados permitidos que encuentra GSA pertenecen al mismo directorio, a través del filtrado de directorios duplicados, el dispositivo GSA los agruparía en un mismo resultado y debería buscar otros 99 resultados permitidos antes de mostrar los resultados de búsqueda.

Capítulo 4 Diseñar una arquitectura para índices de gran escala

En implementaciones con un gran número de documentos, volver a indexar todo el contenido suele llevar varios días o incluso semanas. En estos casos puede ser de gran ayuda pensar en estrategias de diseño de arquitectura para optimizar la manera de indexar los documentos.

Estrategias de rastreo de GSA

Para rastrear contenido web con GSA, hay varias estrategias que permiten mejorar los tiempos de indexación:

- La configuración de **carga de host** puede servir para aumentar el número de hilos que se emplea para rastrear los servidores web y controlar los periodos en los que el rastreo se puede realizar con mayor eficacia.
- Si es posible, deberían usarse varias **rutas de inicio**, ya que esto generaría diferentes hilos de rastreo. Si solo hay una ruta de inicio para una gran cantidad de contenido se ralentizará el rastreo/descubrimiento, a diferencia de cuando se configuran varias rutas de inicio.
- Asegúrese de que la **ubicación de red de GSA** esté lo más cerca u optimizada posible con respecto al servidor web para minimizar así la latencia al rastrear y descargar archivos.
- Si la productividad de rastreo es de vital importancia, puede usar el **rastreo y publicación de resultados distribuido de GSA** para repartir la carga de rastreo en los diferentes dispositivos en la red de DC+S.

Estrategias de transferencia por conector

En implementaciones a gran escala con conectores, se recomienda aplicar una estrategia para transferencias mediante conector para sacar partido a la indexación paralela y tener en cuenta la importancia y las prioridades de los distintos contenidos que se indexan.

Estos son algunos ejemplos de estrategias comunes para indexar repositorios con varios millones de documentos:

- Implemente un conector por cada administrador de conectores y maximice el número de hilos que se usan en el conector (si está disponible, como ocurre con el conector para sistemas de archivos, se suele poder especificar en los archivos de propiedades del conector).
- Configure cada uno de los conectores para que indexe un subconjunto de contenido (p. ej., 3 millones de documentos por conector). Si el contenido no se puede distribuir fácilmente en categorías, se pueden usar los Patrones de Seguir del conector para dividir el contenido por patrones, por ejemplo, dividir los directorios entre los conectores por orden alfabético, como [a-g], [h-m], etc.
- Aumente la carga de host de GSA o implemente varios conectores para incrementar el rendimiento de la transferencia (similar a la solución para aportar mayor disponibilidad descrita anteriormente).

- Si hay contenido prioritario y más importante que otro, asegúrese de que se refleja en la estrategia de transferencia mediante conector. Para controlarlo, aumente la velocidad de transferencia (para garantizar que el contenido se indexa primero) u organice las programaciones de transferencia para que el contenido con prioridad se indexe antes que el resto.