¿Qué es Elasticsearch?

ElasticSearch es un montor de análisis y búsqueda distribuida que facilita la recopilación y agregación de los datos y el almacenamiento en Elasticsearch gracias a Logstash y Beats.

Elasticsearch proporciona búsqueda y análisis casi en tiempo real para todo tipo de datos, ya sean estructurados o no estructurados, Elasticsearch puede almacenarlos e indexarlos de una forma eficiente que posibilite las búsqueda rápidas. También puede recuperar y agregar datos para reconocer tendencias y patrones en los datos.

Elasticsearch ofrece velocidad y flexibilidad para manejar los datos en una gran variedad de casos de uso. Algunos ejemplos pueden ser el agregar un cuadro de búsqueda a una aplicación o sitio web, utilizar machine learning para modelar de forma automática el comportamiento de los datos en tiempo real, almacenar y analizar registros, métricas y datos de eventos de seguridad, etc.

Entrada de información: Documentos e índices

Elasticsearch almacena estructuras de datos complejas que se han serializado como documentos JSON. Cuando existen varios nodos de Elasticsearch en un clúster, los documentos almacenados se distribuyen por todo el clúster y se puede acceder a ellos de inmediato desde cualquier nodo.

Cuando se almacena un documento, se indexa y se puede buscar por completo casi en tiempo real. Elasticsearch utiliza una estructura de datos llamada índice invertido que admite búsquedas de texto completo muy rápidas.

Elasticsearch también tiene la capacidad de no tener esquemas, por lo que los documentos se pueden indexar sin especificar explícitamente cómo manejar cada uno de los diferentes campos que pueden aparecer en un documento gracias al mapeo dinámico.

Salida de información: Buscar y analizar

Gracias a Elasticsearch, además de almacenar y recuperar documentos, se puede acceder a todas las capacidades de búsqueda integradas en la biblioteca del motor de búsqueda Apache Lucene. Con la API REST se puede administrar el clúster e indexar y buscar los datos.

Buscar datos

La API REST de Elasticsearch admiten consultas estructuradas, consultas de texto completo y consultas complejas que combinan las dos. Las consultas de texto completo encuentran todos los documentos que coinciden con el string de consulta y los devuelven ordenados según la relevancia. Se puede realizar búsqueda de frases, similitudes y prefijos, y obtener sugerencias de autocompletado.

Analizar datos

Las agregaciones de Elasticsearch le permiten crear resúmenes complejos de sus datos y obtener información sobre métricas, patrones y tendencias importantes. Permite responder preguntas más específicas y sútiles que

aporten más información a lo que se está buscando. Las agregaciones son muy rápidas, ya que también utilizan las mismas estructuras de datos que la búsqueda, lo que permite ver y analizar datos en tiempo real.

Además, también funcionan con las solicitudes de búsqueda, por lo que se puede buscar documentos, filtrar información y realizar un análisis de los datos al mismo tiempo en una misma solicitud.

Escabilidad y Resiliencia

En Elasticsearch se pueden añadir servidores (nodos) en un clúster para aumentar la capacidad y también distribuye de forma automática la carga de datos y consultas en los nodos disponibles. Elasticsearch balancea los clústeres que tienen varios nodos.

Esto funciona gracias a los fragmentos (shards) de Elasticsearch. Un índice es realmente una agrupación lógica de uno o más fragmentos físicos, donde cada fragmento es un índice autónomo. La distribución de los documentos en un índice en varios fragmentos y esos fragmentos en varios nodos garantiza redundancia, lo que protege contra las fallas de hardware y aumenta la capacidad de consulta.

Hay dos tipos de fragmentos: Primarios y réplicas. Las réplicas son copias de un fragmento primario y protegen contra fallas de hardware y aumentan la capacidad para atender solicitudes de lectura, como buscar o recuperar un documento.

El rendimiento es importante...

Para garantizar un buen rendimiento hay que tener en cuenta el tamaño y cantidad de fragmentos a usar, lo cual va a depender. Aunque la mejor manera de determinar la configuración óptima es realizando pruebas con los datos y consultas propias, en general, es bueno seguir el siguiente punto de partida:

- Tratar de mantener el tamaño del fragmento promedio entre unos pocos GB y algunas decenas del GB.
- Como regla general, la cantidad de fragmentos por GB de espacio de almacenamiento dinámico debe ser inferior a 20.

En caso de desastre

Para que los nodos tengan buenas conexiones entre ellos se colocan en el mismo centro de datos o cerca, pero para mantener una alta disponibilidad se debe evitar cualquier punto único de falla, puesto que en caso de fallo, los servidores localizados en otro lado deben poder tomar el lugar. Para esto se usa la replicación entre clústeres (Cross-Cluster Replication), que permite sincronizar automáticamente los índices de su clúster principal con un clúster remoto secundario que puede servir como copia de seguridad.

Cuidado y alimentación

Algunas herramientas para asegurar, administrar y monitorear los clústeres de Elasticsearch son Kibana (centro de control) y funciones como los resúmenes de datos y la administración del ciclo de vida de los índices para administrar los datos de forma inteligente.

Resumen de ¿What is Elasticsearch?