Motivación ELK

Una estrategia integral de gestión y análisis de registros es de misión crítica, permitiendo a las organizaciones comprender la relación entre los eventos operativos, de seguridad y de gestión del cambio y mantener una comprensión integral de su infraestructura.

Los archivos de registro de los servidores web, las aplicaciones y los sistemas operativos también proporcionan datos valiosos, aunque en diferentes formatos y de manera aleatoria y distribuida.

También los datos que puedan ser extraídos de las redes sociales son buenos insumos para poner en práctica las habilidades de ELK, eso sí se debe tener conocimiento de algunas herramientas que puedan "escudriñar" las redes sociales. En este caso Twitter4J es una buena alternativa.

Los registros son una parte crucial de cualquier sistema porque le dan una idea de lo que un sistema está haciendo así como lo que le sucedió. Prácticamente todos los procesos que se ejecutan en un sistema generan registros de una forma u otra. Estos registros normalmente se escriben en archivos en discos locales. Cuando su sistema crezca a varios hosts, la administración de los registros y el acceso a ellos podría resultar complicado.

La búsqueda de un error particular en cientos de archivos de registro en cientos de servidores es difícil sin herramientas adecuadas. Un enfoque común para este problema es configurar una solución de registro centralizada para que se puedan agregar varios registros en una ubicación central.

Para consolidar, administrar y analizar eficazmente estos diferentes registros, muchos clientes optan por implementar soluciones de registro centralizadas utilizando Elasticsearch, Logstash y Kibana, popularmente conocido como ELK Stack.

¿Cuáles son las posibles aplicaciones para Elasticsearch?

El caso de uso clásico para Elasticsearch es la búsqueda de texto libre. Además, sin embargo, también puede ser utilizado en la sección Analytics para evaluar los grandes

conjuntos de datos en "tiempo real". Estos son algunos ejemplos de la utilización de Elasticsearch.

i. Si desea ampliar su web o aplicación móvil con una función de búsqueda que proporcione al usuario características, tales como, autocompletado o clasificación de los resultados de búsqueda en función de su relevancia.
ii. Si desea implementar una aplicación móvil que localiza las empresas más cercanas utilizando la ubicación actual del usuario, las muestra en un mapa u ordenados por la distancia recorrida. En tal caso, Elasticsearch ofrece muchas maneras de implementar una búsqueda basada en la localización. Por lo tanto, los datos pueden ser enriquecidos, por ejemplo, con coordenadas y resultados de búsqueda se filtran por la distancia a una ubicación particular durante la indexación.
iii. Si desea crear informes bajo un determinado Lenguaje de Dominio Específico (DSL) bajo el concepto que entrega el DSL de Elasticsearch en la consulta.
iv. Si desea evaluar los datos de transacciones y analizarla para hacer ciertas tendencias y desviaciones visible, o desea realizar estadísticas y consultas ad-hoc de grandes conjuntos de datos para satisfacer sus necesidades en el ámbito de la Inteligencia de Negocio(BI) o bien Advanced Analytics. También puede utilizar Logstash como herramienta para la transformación y proporcionar sus datos en Elasticsearch. También poner en Kibana los datos en una interfaz gráfica e interactiva para visualizar y evaluar, así como realizar sobre la agregación Elasticsearch, que permite consultas complejas de BI.

La configuración de la pila ELK tiene tres componentes principales:

Elasticsearch: Se utiliza para almacenar todos los registros de aplicación y de supervisión.

Logstash: El componente de servidor que procesa registros de entrada.

Kibana: Una interfaz web para buscar y visualizar registros.

Aplicaciones: Temas para ser abordados como proyectos transversales.

Un ejemplo concreto: se tiene una aerolínea y queremos saber cómo es la experiencia de los viajeros y cuál es el perfil de viajeros que más viaja. Una alternativa es usar la API de Twitter en donde podríamos realizar una búsqueda de los aeropuertos a los que nuestra compañía vuela, que lleven la palabra vuelo, de tal manera que lleguemos a sacar datos sobre cuántos usuarios se han quejado en esta red social, así como sobre la cancelación o retraso de sus vuelos y cuál es el promedio, por ejemplo, de seguidores, de esos usuarios. De esta manera, sean o no clientes de nuestra aerolínea, sabremos qué aeropuertos suelen tener más problemas y cuáles son los tipos de usuarios que usan esta red social para dar a conocer la incidencia o, inclusive, como medio de atención al cliente.

Con este mismo ejemplo, si lo consideramos en un ámbito universitario, poder saber los comentarios que se realizan en el contexto académico de tal o cual profesor en el ámbito de su docencia, o cuantas veces ha sido citado, etc. O bien, los comentarios que se realizan sobre las competencias de tal o cual profesional, por ejemplo, algún médico.

Otro ejemplo práctico del uso de la API en un sector diferente podría ser para una compañía de telecomunicaciones, gracias a las búsquedas en la red social, podemos conocer cuáles son las razones más habituales para darse de baja y, de nuevo, cuál es el perfil de usuario más satisfecho y menos satisfecho con nuestros servicios.

Con este mismo ejemplo, si lo consideramos en un ámbito universitario, poder saber los comentarios que se realizan en el contexto de los servicios administrativos que se ofrecen, etc.

En el caso de un banco, por ejemplo, se pueden realizar búsquedas sobre cajeros automáticos y específicamente sobre robos en cajeros automáticos o fallos en estas máquinas. Luego, para tomar medidas, tales como ampliar la seguridad en los lugares más conflictivos es una buena oportunidad. O también de realizar un filtrado con búsquedas más específicas como: “crédito, banco, satisfecho o buen”, “cliente, banco, fiel o aconsejable” para averiguar qué servicios o con qué están más contentos sus clientes con cuenta en Twitter.

1. **Bueno para ello, previamente, deberán ¿ Cómo obtener datos desde una cuenta twitter usando Java.?**

https://www.ibm.com/developerworks/library/j-use-elasticsearch-java-apps/index.html

2. **Indexando Twitter con Logstash y Elasticsearch**

http://david.pilato.fr/blog/2015/06/01/indexing-twitter-with-logstash-and-elasticsearch/

3. **Usando Kibana y Elasticsearch para examinar Twitter Trends.**

https://qbox.io/blog/using-kibana-and-elasticsearch-to-examine-twitter-trends

4. **Twitter API, ElasticSearch y Kibana – Analizando las redes sociales… ¿Quién hace el mejor pronóstico respecto a las candidaturas presidenciales, con llegadas de 1 a 8.?**

- **https://imasters.com.br/apis/apis-twitter/twitter-api-elasticsearch-e-kibana-analisando-a-rede-social/?trace=1519021197&source=single**
- **https://logz.io/blog/analyzing-twitter-elk-stack/**

5. **Detectando Terremotos con Twitter y Elasticsearch**

https://thenewstack.io/detecting-earthquakes-twitter-elasticsearch/

Referencias:

https://www.elastic.co/webinars/introduction-elk-stack

https://www.elastic.co/products

Elasticsearch, Una guía completa..

https://books.google.cl/books?id=v08oDwAAQBAJ&pg=PA53&lpg=PA53&dq=twitter+api+and+elasticsearch&source=bl&ots=3_raYUiKnV&sig=B2Q1XihA5qwu0tlB8oXMMICIEAc&hl=es&sa=X&ved=0ahUKEwihksv5_u3VAhWBS5AKHc5DBRA4FBDoAQg-MAQ#v=onepage&q=twitter%20api%20and%20elasticsearch&f=false

Rumor Detection and Prediction on Social Media

Wei Wang (Team Leader) and Weisheng Zhong