ELASTIC STACK

MODALIDAD DE OPCIÓN DE GRADO, TRABAJO DE FORMACIÓN EMPRESARIAL
OPTION MODE OF DEGREE, BUSINESS TRAINING WORK

DANIEL FELIPE FAJARDO PENAGOS

EDGAR RUIZ DORANTES

INGENIERÍA DE SISTEMAS

CIENCIAS E INGENIERÍA

FUNDACIÓN UNIVERSIDAD JORGE TADEO LOZANO

2020-06-19

RESUMEN EJECUTIVO

En Accenture surge la necesidad de detectar en tiempo real las incidencias y/o inconvenientes que surjan en la operación de monitorización de carga de datos Big Data, dando respuesta en tiempo real a estas eventualidades de error, permitiendo escalabilidad en el sistema.

Para solventar esta necesidad se emprenderá en la implementación del software Elastic Stack que provee las funcionalidades de motor de búsqueda analítica, procesamiento, transformación y almacenamiento de datos, y además permite visualizar los datos en cuadros y gráficos dinámicos. De esa forma se va se reacciona y adapta sin perder calidad en el procesamiento de grandes volúmenes de información obtenido grandes rendimientos en el procesamiento de datos dentro de la operación.

EXECUTIVE OVERVIEW

At Accenture the need arises to detect in real-time the incidents and/or inconveniences that arise in the Big Data load monitoring operation, giving a real-time response to these eventualities of error, allowing scalability in the system.

To solve this need, Elastic Stack software will be implemented, which provides the analytical search engine, data processing, transformation, and storage functionalities, and allows the visualization of data in dynamic charts and graphs. In this way, it will react and adapt without losing quality in the processing of large volumes of information, obtaining high performance in data processing within the operation.

OBJETIVOS

## Objetivo General

Adecuar e implementar una plataforma de analítica y monitoreo que permita dar seguimiento a los registros en tiempo real en el área de AIP de la empresa, con el fin de optimizar el tiempo de monitoreo y mediante esa optimización realizar mejoras de proceso necesarias.

## Objetivos Específicos

* Documentar e investigar sobre diferentes aplicaciones y/o herramientas de monitoreo de registros similares de la propuesta por el trabajo de formación empresarial
* Definir y ejecutar las actividades requeridas por la herramienta Elastic Stack para el correcto funcionamiento en las diferentes máquinas virtuales
* Convocar reuniones con el cliente directo para su validar los requerimientos solventados y su posterior aprobación
* Generar test cases con el fin de realizar una carga inicial de datos donde se pueda comprobar el procesamiento correcto de registros a través de las capas de herramientas dentro de Elastic Stack, asegurando así la calidad con la que los registros son procesados
* Proponer mejoras de procesos de negocio con base en los resultados obtenidos por medio del software Elastic Stack

DESCRIPCIÓN DEL PROCESO

A continuación, se describe de manera clara, secuencial y coherente las actividades que se desempeñaron en la práctica empresarial.

A lo largo de mi experiencia por Accenture tuve que manejar y monitorear grandes volúmenes de datos por medio de procesos que involucran el movimiento de datos desde múltiples fuentes de datos, transformarlos, limpiarlos y cargarlos en otra base de datos, proceso comúnmente denominado ETL. Estos procesos se realizaban con herramientas líder en integración, calidad y administración de datos

Una de las piezas claves para logar este proceso de ETL diariamente con eficacia, fue el afincamiento a las metodologías ágiles SCRUM en la práctica y la teoría. En Accenture, cada día mi equipo de trabajo nos preguntábamos sobre ¿qué se va a ser a lo largo del día?, ¿qué se hizo con respecto al último sprint?, y si ¿Existe algún impedimento para realizar los objetivos del día? Todo esto se realizó con el fin de maximizar la satisfacción del cliente.

Hubo actividades, que, en mi opción, me ayudaron a crecer en habilidades como en la escucha y comunicación eficaz, gestión y control de el tiempo. Esto se logró gracias a reuniones directas con el cliente para la medición de progresos, construcción y/o mantenimiento de las actividades, además del control en la trayectoria que va tomando la visión del producto.

A continuación, se evidencia la labor realizado a lo largo de mi primera etapa en Accenture.



Ilustración 1. Fuente Propia

Además, se realizaron procesos de analítica descriptiva por medio de hermanitas de Python que permiten la manipulación de datos en DataFrames, alineación de datos y manejo integrado de datos faltantes, POO y gráficos.

DESARROLLO

Durante el desarrollo de mis prácticas se llevaron a cabo actividades, en la que requirió gran parte de mi participación, de ahora en adelante llamada iniciativa. En el cuál se desarrolló una arquitectura Big Data para la ingestión, enriquecimiento, almacenamiento, monitorización, análisis y visualización de datos.

El alcance de esta iniciativa es poder identificar las dependencias a nivel de networking que se necesitan para conectar los servidores generadores de logs, además poder monitorizar de forma adecuada la arquitectura utilizada por los servidores y herramientas de software.

El primer paso para desarrollar la arquitectura necesaria por Elastic Stack en un ambiente productivo es entender y conocer los requerimientos necesarios para el despliegue de este, por lo cual se desarrolla esta arquitectura de alto nivel para su entendimiento. En este primer desarrollo fueron necesarios conocimientos adquiridos en asignaturas de arquitectura de software y arquitectura empresarial



Una vez entendiendo el diseño de arquitectura se procede a realizar el despliegue en un ambiente de pruebas. Por lo que, una vez realizado, se procede a la instalación y configuración de las herramientas de Elastic Stack empezando con Filebeats, una herramienta que funciona como un agente que recoge los logs de sistemas y software específico.

Para esta etapa fue necesario la configuración de archivo filebeat.yml, archivo de configuración predeterminado. Es el encargado de supervisar los archivos de registro o las ubicaciones que especifique, recopila eventos de registro y los reenvía a Elasticsearch o Logstash para su indexación (Elastic, 2020). Aquí está una muestra de la sección del archivo.



El siguiente paso es la configuración entre los componentes Filebeats, Logstash y Elasticsearch. Lo cuál se logra indicando la IP del servidor y el puerto por el cual van a estar escuchando las herramientas del Stack de la siguiente forma.





La configuración de la herramienta logstash, se hace por medio de parseo de logs y de pipelines que ayudan a la automatización en la ejecución de la herramienta. Para el parseo de logs se utilizan expresiones regulares que van a ayudar a captar los patrones de logs de los sistemas como se evidencia a continuación.



En la configuración de pipelines, se separan las ejecuciones de procesamiento en diferentes archivos que tienen como entrada logs de un sistema determinado. Cada uno de estos archivos se va a encargar de procesar diferentes fuentes de logs. Para esta configuración se especificó las rutas de estos archivos en el archivo pipelines.yml para que al levantar Logstash como servicio, encendiera con estos archivos de configuración por defecto. La evidencia del correcto funcionamiento de esta ejecución se evidencia a continuación.





Un reto importante que surge durante el desarrollo de la iniciativa es la identificación de los archivos de log, esto porque todos los registros son almacenados en una misma carpeta. El solventar este reto es importante ya que es requerimiento funcional del cliente poder segmentar los registros de log por cada fuente se aplicación y/o sistema. La solución a este reto es parsear el nombre del archivo y extraer el nombre del log, con el cual, por medio de un script Ruby, se extrae el id.

Para las configuraciones de Elasticsearch se realizan un grupo de querys, que son unas consultas mediante el lenguaje Domain Specific Language (DLS). Se realizó una primera query definida por términos en el que deberá traer los procesos definidos en la variable, al cual se le incluyen 4 agregaciones que permitirá agrupar información por campos definidos. La agregación process\_name permite agrupar los datos por nombre de proceso en los que se ejecutan los mappings, la agregación lvl\_log permite agrupar los datos por el nivel de log con los que se encuentra, la agregación name\_mpn permite agrupar los nombres de mapping que se ejecutaron, la agregación app\_name permite agrupar los datos por el nombre de la app lanzada. Se realizó otra query definida por términos y por wildcard en el que la información se verá segmentada por el id de la ejecución de los workflows y por el tipo definido en el campo. Se realizan 3 agregaciones, la primera de ellas es sucessfull que permite agrupar los datos por los procesos que se completan con normalidad dentro del workflow, la agregación section permitirá agrupar los datos por el nombre de las ejecuciones. A continuación, se ilustra la evidencia de funcionamiento de las querys.



Finalmente se realiza la configuración de la herramienta de visualización de los datos, Kibana. Para esta configuración se llevó a cabo 3 diferentes dashboards que ilustran información como la cantidad de log que se completan con normalidad, alertas o deficientes en el tiempo; detalle de log que falló, estatus de log, opción de filtro de la información. A continuación, se ilustra uno de los 3 dashboards generados para la empresa Accenture.



CONCLUSIONES

* En estas prácticas laborales aprendí a aplicar los conocimientos que adquirí a lo largo de mi carrera profesional en la universidad Jorge Tadeo Lozano.
* Una de las experiencias más retadoras a mi experiencia, siendo la más provechosa, fue al momento de desenvolverme en mis habilidades de comunicación, donde tuve la necesidad de comunicarme de forma efectiva, directa con el cliente
* En el desarrollo e implementación de la herramienta Elastic Stack para la monitorización y analítica de grandes volúmenes de datos, descubrí un verdadero gusto por la analítica de datos, algo que desde la universidad iba creciendo en mí y en el área profesional se afianzó.
* Durante la práctica hubo momentos en donde la solución a un problema no era tan obvia, además con entregas con tiempo limitado, por lo que se me obligó a tener una capacidad de resolución de problemas y gestión del estrés altos para alcanzar mis objetivos en ese momento.