

Proyecto FINAL

ESCUELA POLITÉCNICA NACIONAL **ESCUELA DE FORMACIÓN DE TECNÓLOGOS**

MAURICIO RODRÍGUEZ
JOSUE GARRIDO
XAVIER JARAMILLO
mauricio.rodriguez01@epn.edu.ec
francisco.jaramillo01@epn.edu.ec
josue.garrido@epn.edu.ec

Antes de instalar cualquier programa se debe tener en cuenta lo siguiente:

- El espacio en disco
- El sistema operativo de 32 o 64 bits
- La Capacidad de RAM
- Y Versión de tu maquina

En esta ocasión procederemos con el entendimiento de como instalar programas dedicados en este caso de programación. Con enfoque a minería de Base de Datos Multinacionales y la debida utilización de herramientas dedicadas al mismo.

I. DEFINICION DEL CASO DE ESTUDIO

El caso de estudio que hemos de realizar con la recopilación de información utilizaremos filtros de palabras para poder subdividir los temas presentados en los ítems, para posteriormente poder analizarlos mediante herramientas antes vistas en el semestre como por ejemplo: KIBANA, GOOGLE CHART, con esta información vamos a visualizar de manera gráfica, sencilla y fácil de entender.

II. OBJETIVOS GENERALES Y ESPECIFICOS

Objetivo General: Identificar los datos concentrados en cada una de las temáticas mediante una interfaz gráfica.

Objetivo Específico: Analizar la mejor opción para el almacenamiento masivo de datos.

III. DESCRIPCION DEL EQUIPO DE TRABAJO Y ACTIVIDADES DE TRABAJO RESALIZADAS POR CADA UNO.

Nuestro equipo de trabajo siendo analistas en sistemas de la información debemos tener un claro y amplio entendimiento del uso de datos ya que estos están inmersos en diferentes campos ya sean profesionales o no.

Actividades:

Josué Garrido:

- Indagación de base de datos SQL
- Selección de la base de datos SQL
- Instalación de la base de datos SQL
- Realización de la minería de datos en las plataformas twitter y Facebook

- Recopilación de datos
- Creación del dashboard en kibana
- Presentación de datos

Xavier Jaramillo:

- Indagación de base de datos No SQL
- Elección de base de datos No SQL
- Descarga de los softwares (couchdb y mongo db)
- Instalación
- Realización de la minería de datos en couchdb
- Creación del dashboard en kibana
- Presentación de datos

Mauricio Rodríguez:

- Instalación de base de datos No SQL (couchdb y mongo db)
- Realización de la minería de datos en Mongoddb
- Recopilación de datos
- Creación del dashboard en kibana
- Presentación de datos

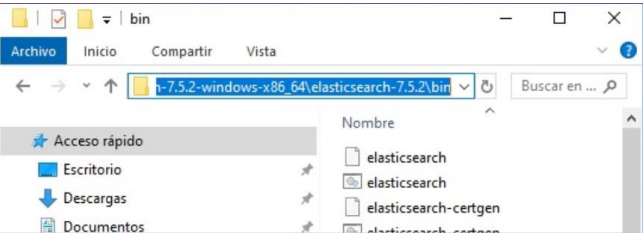
IV. CRONOGRAMA DE ACTIVIDADES



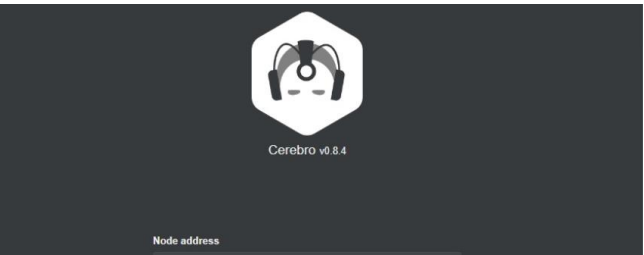
V. ASIGNACION DE ACTIVIDADES A CADA MIEMBRO DEL EQUIPO

VI. RECURSOS Y HERRAMIENTAS USADAS

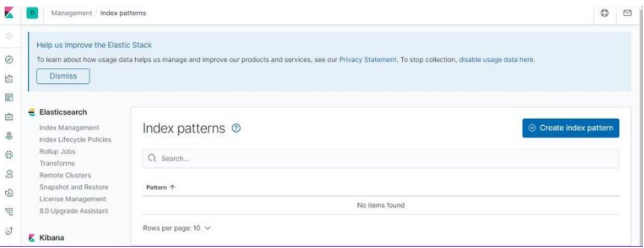
Elasticsearch



Cerebro



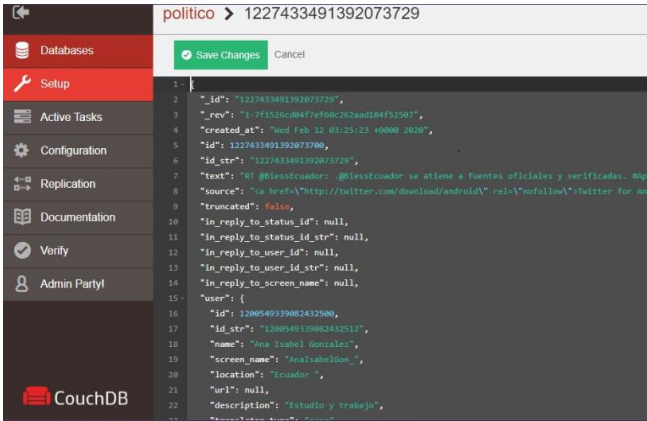
Kibana



Databases			
Name	Size	# of Docs	Actions
eventos	10.4 MB	3955	
politico	55.8 MB	19132	
quejas	24.6 MB	8640	
trafico	16.2 MB	5388	
twitter	30.1 MB	10729	

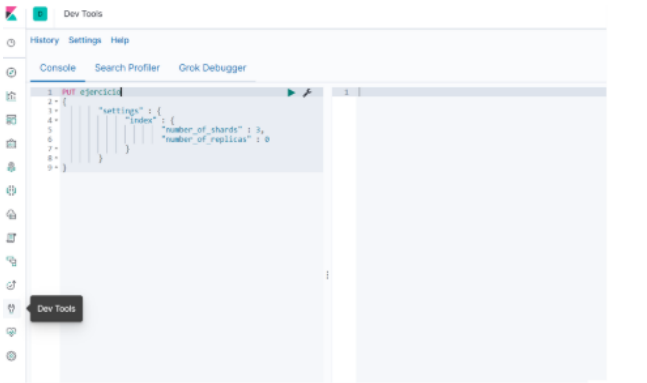
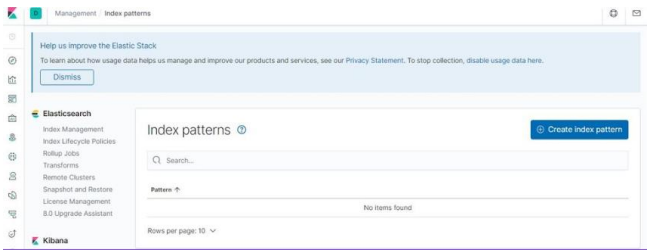
IX. ANALISIS DE INFORMACION

En los diferentes casos y diferentes items los cuales obtuvimos podemos observar que al momento de visualizar y minar los datos de algunos de los temas no era factible ya que en ocasiones el y los filtros ya que palabras clave en los determinados temas existen conflictos al momento de minar

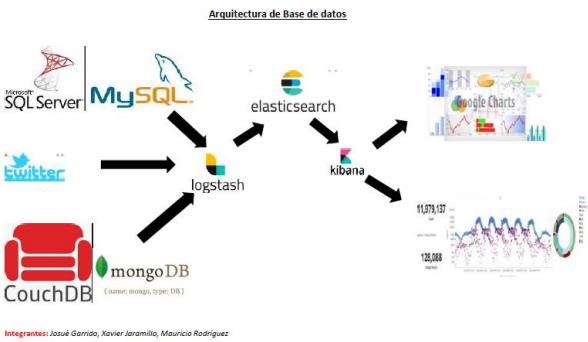


X. VISUALIZACION DE LA INFORMACION

Como podemos observar lo primero es minar los datos y demás en kibana para lo cual levantamos lo siguiente



VII. ARQUITECTURA DE LA SOLUCION



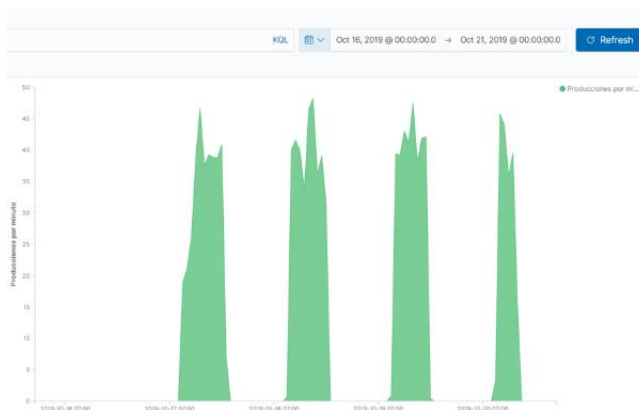
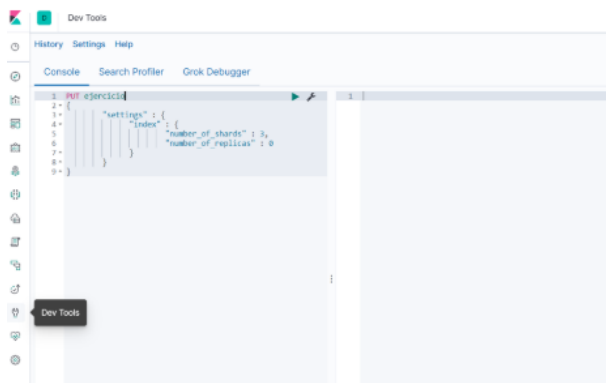
VIII. EXTRACCION DE DATOS

Para esto utilizamos la cosecha.py y utilizamos los diferentes filtros a continuación los mostramos

En este caso podemos observar los datos recopilados de la pasteurizadora quito con mas de 5 millones de datos los cuales nos muestra la producción por minuto



XI. RESULTADOS OBTENIDOS



XII. CONCLUSIONES Y RECOMENDACIONES

Como principales conclusiones podemos decir que la extracción de datos es verdaderamente fácil ya que con los API de Twitter y Facebook podemos realizarlos de mejor manera y rápidamente y con ella la manipulación de los datos con Kibana es verdaderamente eficiente.

Recomendamos primero transformar los datos extraídos con Elasticsearch a un archivo de Excel para poder visualizarlos de mejor manera.

XIII. DESAFÍOS Y PROBLEMAS ENCONTRADOS

- Desafío más significativo fue la extracción de datos en los cuales el filtro conlleva conflictos de escritura
- Uno de los problemas fue con la base de datos MongoDB ya que los APIs de Twitter nos establecen problemas tales como bloqueos y demás.

XIV. LINKS DE GITHUB DEL PROYECTO

https://github.com/JosueGarrido/ProyectoFinal_Base_De_Datos_Multidimensionales_DataLake/upload/master

XV. BIBLIOGRAFIA

[1] "Cómo instalar MongoDB en Windows, Linux y Mac", Platzi, 2020. [Online]. Available: <https://platzi.com/blog/como-instalar-mongodb-en-window-linux-y-mac/>. [Accessed: 12- Feb- 2020].

[2] "Primeros pasos con CouchDB - Adictos al trabajo", Adictos al trabajo, 2020. [Online]. Available: <https://www.adictosaltrabajo.com/2018/11/27/primeros-pasos-con-couchdb/>. [Accessed: 12- Feb- 2020].

[3] "{metadataController.pageTitle}", Subscription. packtpub.com, 2020. [Online]. Available: https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781786465580/12/ch12lvl1sec144/installing-and-using-cerebro. [Accessed: 13- Feb- 2020].