

# Arquitectura Data Lake

## Informe (Septiembre de 2021)

Quinatoa Johan, Quisaguano Bryan, Rea Jhosef, Sanchez Antony

### I. OBJETIVOS

#### Objetivo General

- Diseñar una arquitectura Data lake.

#### Objetivos Específicos

- Recopilar datos de fuentes de internet referente a los temas dados.
- Almacenar en bases de datos SQL y NoSQL la información recopilada.
- Utilizar elasticsearch como concentrador de datos.
- Realizar un caso de estudio para cada temática.

### II. DEFINICIÓN DEL CASO DE ESTUDIO

Pulso político en 20 ciudades del país, se pretende recolectar toda la información posible de los temas políticos que más han sonado en este año, en especial luego de las elecciones presidenciales y de delegados de asambleístas, esto servirá para calcular el nivel de aceptación o conformidad que tienen los habitantes por sus nuevos gobernantes.

Pulso político en provincias del Ecuador. Debido a que en este año sucedieron las elecciones tanto presidenciales como de representantes asambleístas, los datos recopilados y analizados, permiten evidenciar cuán diferente fue la decisión de los habitantes al momento de acudir a las votaciones Y el candidato o lista mayormente votadas por provincias.

Juegos en línea. No cabe duda de que en época de pandemia los videojuegos han sido para muchas personas un medio de distracción y relajación dentro del confinamiento, por lo cual los videojuegos se han convertido aún más en el principal medio de entretenimiento, con esto en mente se pretende analizar los datos que se obtengan de países, para así saber: los juegos que son más populares dentro de una serie de países. Además, se obtendrá el país que tiene una mayor producción de videojuegos exitosos.

Python. Ya que hoy en día el lenguaje de programación Python es uno de los más populares en el mundo, se pretende obtener datos para así saber que tan grande es la comunidad de América Latina y que tanto este tema es mencionado dentro de las redes sociales.

Noticias mundiales. Las noticias que causan revuelo internacional son el pan de cada día en emisoras tanto de radio como de televisión, empleando distintas herramientas de pretende saber qué tanto revuelo puede causar diversas noticias que sean catalogadas como internacionales dentro de internet específicamente redes sociales.

### III. CRONOGRAMA DE ACTIVIDADES

Las actividades se realizaron siguiendo la siguiente planificación descrita a continuación en las siguientes tablas:

Cronograma		L	M
Actividades	RESPONSABLE	6 septiembre	8 septiembre
Creación de un repositorio en GitHub para almacenar las bases de datos	Grupo		
Definir casos de estudio para: Pulso Político, Juegos en línea por países, Tema libre, Eventos o noticias mundiales	Grupo		
Creación de la arquitectura de cada BD	Grupo		
División en grupos para cada subtema	Grupo		
Planificación de actividades	Grupo		
Creación de la arquitectura de conexión	Grupo		

Pulso político	Antony Sanchez, Jhosef Rea		
Recopilación de datos para Pulso Político a través de CSV	Antony Sanchez, Jhosef Rea		
Creación de Base de Datos SQL	Antony Sanchez, Jhosef Rea		
Transformación de datos para insertar en una base de datos SQL	Antony Sanchez, Jhosef Rea		
Inserción de datos en BD	Antony Sanchez, Jhosef Rea		

Cronograma		M	J
Actividades	RESPONSABLE	8 septiembre	9 septiembre
Juegos en línea	Johan Quinatoa, Jhosef Rea		
Recopilación de datos para Juegos en línea por país a través de web scraping	Johan Quinatoa, Jhosef Rea		
Creación de Base de Datos SQL	Johan Quinatoa, Jhosef Rea		
Transformación de datos para insertar en una base de datos SQL	Johan Quinatoa, Jhosef Rea		
Inserción de datos en BD	Johan Quinatoa, Jhosef Rea		
Tema libre: python	Bryan Quisaguano, Johan Quinatoa		
Creación de Base de Datos NOSQL - MongoDB	Bryan Quisaguano, Johan Quinatoa		
Recopilación de datos para Tema libre a través de twitter	Bryan Quisaguano, Johan Quinatoa		
Transformación de datos para insertar en una base de datos nosql	Bryan Quisaguano, Johan Quinatoa		
Inserción de datos en BD	Bryan Quisaguano, Johan Quinatoa		

Cronograma		S	L
Actividades	RESPONSABLE	11 septiem bre	13 septiem bre

Eventos o noticias mundiales	Antony Sanchez, Bryan Quisaguano		
Creación de Base de Datos NOSQL - Couch	Antony Sanchez, Bryan Quisaguano		
Recopilación de datos para Eventos o noticias mundiales a través de tik tok	Antony Sanchez, Bryan Quisaguano		
Transformación de datos para insertar en una base de datos nosql	Antony Sanchez, Bryan Quisaguano		
Inserción de datos en BD	Antony Sanchez, Bryan Quisaguano		
Crear indexación de las BD	Grupo		
Conectar los BD a Kibana con cerebro	Grupo		
Creación del informe	Grupo		

Cronograma		M	M
Actividades	RESPONSABLE	14 septiembre	15 septiembre
Creación de la presentación	Grupo		
Creación del video	Grupo		

Cronograma		J
Actividades	RESPONSABLE	16 septiembre
Subida del proyecto	Grupo	

#### IV. HERRAMIENTAS UTILIZADAS

- Apache CouchDB

Es un gestor de bases de datos de código abierto, cuyo foco está puesto en la facilidad de su uso y en ser "una base de datos que asume la web de manera completa". Se trata de una base de datos NoSQL que emplea JSON para almacenar los datos.

- MongoDB

Es un sistema de base de datos NoSQL, orientado a documentos y de código abierto. En lugar de guardar los datos en tablas, tal y como se hace en las bases de datos

relacionales.

- MongoDB Atlas

es un servicio de Cloud Database (o Base de Datos en la Nube), que te permite crear y administrar tu BBDD Mongo además que no solo está orientado a ser accesible desde el navegador, sino que, fue desarrollado con el objetivo de aliviar el trabajo de los desarrolladores, al quitarles la necesidad de instalar y administrar entornos de BBDD, los que a veces pueden ser lentos y engorrosos.

- Python

Es un lenguaje de programación de alto nivel, orientado a objetos, con una semántica dinámica integrada, principalmente para el desarrollo web y de aplicaciones informáticas.

Lo utilizaremos específicamente para realizar scripts que ayudan a la recolección de datos en nuestras bases de datos

- Elastic Search

Es un servidor de búsqueda basado en Lucene. Provee un motor de búsqueda de texto completo, distribuido y con capacidad de multitenencia con una interfaz web RESTful y con documentos JSON. Elasticsearch está desarrollado en Java y está publicado como código abierto bajo las condiciones de la licencia Apache.

- Kibana

Es una herramienta open-source perteneciente a Elastic, que nos permite visualizar y explorar datos que se encuentran indexados en Elasticsearch, es decir, un plugin de Elasticsearch.

- PhPmyAdmin

Es una herramienta escrita en PHP con la intención de manejar la administración de MySQL a través de páginas web, utilizando un navegador web. Actualmente puede crear y eliminar bases de datos, crear, eliminar y alterar tablas, borrar, editar y añadir campos, ejecutar cualquier sentencia SQL, administrar claves en campos, administrar privilegios, exportar datos en varios formatos.

- SQL server

Es un sistema de gestión de bases de datos relacionales (RDBMS) que admite una amplia variedad de aplicaciones de procesamiento de transacciones, inteligencia empresarial y análisis en entornos informáticos corporativos. Microsoft SQL Server es una de las tres tecnologías de bases de datos líderes del mercado, junto con Oracle Database y DB2 de IBM.

- Facebook

Es la principal red social que existe en el mundo. Una red de vínculos virtuales, cuyo principal objetivo es dar un soporte para producir y compartir contenidos. Llegó para ampliar las posibilidades de relación social y causó una revolución sensible en el mundo de las comunicaciones.

Para el proyecto, se utiliza Facebook para recolectar información de páginas para usar esa información para el análisis de datos.

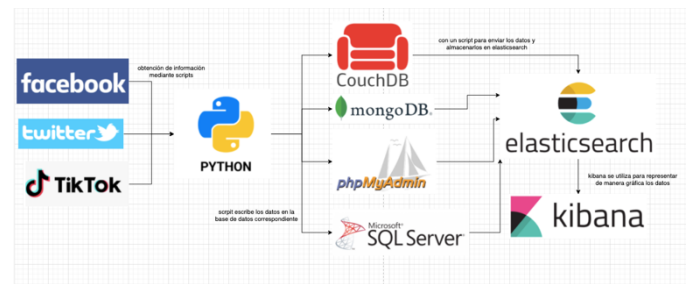
- Twitter

es una red social gratuita que te permite escribir mensajes de forma rápida y sencilla. Cualquier persona puede crear una cuenta o perfil para compartir sus opiniones, gustos e intereses, pero también, para estar informada sobre todo lo que acontece en el mundo, ya que la principal característica de Twitter es que allí encuentras información en tiempo real.

Para el proyecto, se utiliza Twitter para recolectar información de tweets para usar esa información para el análisis de datos.

## V. ARQUITECTURA UTILIZADA

La arquitectura con la que se desarrolló el Data Lake es la siguiente:



## VI. EXTRACCIÓN DE DATOS

Se extrajo datos de las fuentes de Twitter, Facebook, Kaggle, Tik Tok y de ciertas fuentes web en base a los temas propuestos. Fueron extraídos mediante los scripts proporcionados en clase (Facebook, Twitter). También se obtuvieron datasets desde la página de Kaggle y se realizó un web scraper a Tik tok y a páginas web.

A continuación, se muestra una breve explicación de cómo se extrajeron los datos e ingresados a una base de datos en mongodb, en este caso el tópico fue sobre Python.

Reemplazar esta línea con su número de documento de identificación (hacer doble clic aquí para editar) < 4

```

2 from pymongo import MongoClient, collection
3 from facebook_scraper import get_posts
4 import json
5 import time
6
7 myclient = MongoClient("mongodb://localhost")
8 db = myclient['pythonEcuador']
9 col = db['dataFace']
10
11 i=1
12 for post in get_posts('pythonEcuador', pages=25, extra_info=True):
13     print(i)
14     i=i+1

```

## Fig. Script de Facebook a Mongo

[illegible]

Fig. WebScraping

```
@(shell).1.4
> db.dataFace.count()
84
```

Fig. colección creada  
Twitter

```
~ tiktok-scraper user zippycode -t csv
```

CSV path: /home/jhosefjrdbz/zippycode\_1631485870858.csv

Fig. TikTok scraper

```

20 def on_data(self, data):
21     dictTweet = json.loads(data)
22     try:
23
24         dictTweet["_id"] = str(dictTweet['id'])
25         doc = col.insert_one(dictTweet)
26         print ("SAVED" + str(doc) + ">" + str(data))
27     except:
28         print ("Already exists")
29         pass
30     return True
31
32 def on_error(self, status):

```

PROBLEMS    DEBUG CONSOLE    OUTPUT    TERMINAL

```

taScience", "indices": [111, 123], {"text": "Python", "indices": [124, 1
"text": "CodeVsnvie", "indices": [148, 159], {"text": "100DaysOfCode",
", [175, 188], {"text": "BlackTechTwitter", "indices": [189, 206], {"te

```

Fig. colección creada  
Twitter

```
> use pythonEcTwitter
switched to db pythonEcTwitter
> db.dataTwitter.count()
355
```

Fig. colección creada  
Twitter

```
> use pythonTikTok
```

Fig. Base de datos creada TikTok

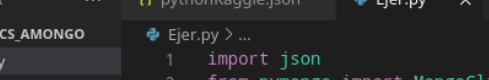
```

> db.dataTtok.find().pretty()
{
  "_id" : ObjectId("613ce8b3c8b3278dd8b366c45"),
  "id" : NumberLong("7007079789511886000"),
  "secretID" : NumberLong("7007079789511886000"),
  "text" : "Speed game dev #speed #speedrun #gaming #gamedev #tech",
  "createTime" : 1631462898,
  "authorMeta.id" : NumberLong("6945122548014331000"),
  "authorMeta.secidId" : "M5W4tLjABAAAYaqe-pRvkpbvtor8oFv-rSAAYk4a-PsMkUmyEq0r_soQlVrLrLz0UHFyVeir-UPcd",
  "authorMeta.name" : "zippycode",
  "authorMeta.nickName" : "ZippyCode",
  "authorMeta.verified" : false,
  "authorMeta.avatar" : "https://p16-sign-va.tiktokcdn.com/tos-maliva-avt-0068/167e85df026893b0f0184e616fb3df-c5_1080x1080.jpeg?x-expires=1631570400&x-signature=Y5Wrn3gr15vMgUXDKY14Goy10u3D",
  "authorMeta.following" : 34,
  "authorMeta.fans" : 305800,
  "authorMeta.heart" : 3000000,
  "authorMeta.video" : 297,
  "authorMeta.digg" : 383,
  "musicMeta.musicId" : NumberLong("6927278997412547000"),
  "musicMeta.musicName" : "The Search - Edit",
  "musicMeta.musicAuthor" : "NfP",
  "musicMeta.musicOriginal" : false,
  "musicMeta.musicAlbum" : "The Search - Edit",
  "musicMeta.playUrl" : "https://s1615-ies-music-va.tiktokcdn.com/obj/tos-useast2a-ve-2774/5f4f0d6502894cadb79274d36fe872",
  "musicMeta.coverThumb" : "https://p77-va.tiktokcdn.com/img/tos-useast2a-ve-2774/dad65dc2a86247b6a944f861ecfac0e3-c5_100x100.jpeg",
  "musicMeta.coverMedium" : "https://p77-va.tiktokcdn.com/img/tos-useast2a-ve-2774/dad65dc2a8624a7b

```

Fig. Colección sobre los datos extraídos deTikTok

A continuación, se presenta el dataset descargado de Kaggle y subido mediante script a MongoDB.



The screenshot shows a code editor interface. On the left, the 'EXPLORER' panel displays a file tree with the following items: `pythonKaggle.json`, `Ejer.py` (selected), `tikTok.json`, and a folder named `SUBIRDOS_AMONGO`. The main editor area shows the content of `Ejer.py`, which contains the following Python code:

```
Ejer.py > ...
1 import json
2 from pymongo import MongoClient
3
4
```

At the bottom, the 'TERMINAL' panel shows the command to run the script:

```
/usr/bin/python3 /home/jhosefjrdbz/PROGRAMAR/~ /usr/bin/python3
irDocs_AMongo/Ejer.py
```

The terminal output shows the current directory path: `~/PROG/P/Análisis/SubirDocs_AMongo`.

Fig. Kaggle

Reemplazar esta línea con su número de documento de identificación (hacer doble clic aquí para editar) < 5

```
> use pythonKaggle
switched to db pythonKaggle
> db.dataKaggle.find().pretty()
```

Fig. Base de datos y Colección creadas.

```
{
  "_id" : ObjectId("613e9183c547cd3ee084de4b"),
  "Week" : 11,
  "This_Week" : 19,
  "Last_Week" : 19,
  "End_2019" : 20,
  "Name" : "Louis Oosthuizen",
  "Average_Points" : 4.3328,
  "Total_Points" : 199.31,
  "Events_Played_Divisor" : 46,
  "Points_Lost_2020" : -33.33,
  "Points_Gained_2020" : 35.49,
  "Events_Played_Actual" : 46,
  "Pro/Am" : "Pro"
}
{
  "_id" : ObjectId("613e9183c547cd3ee084de4c"),
  "Week" : 11,
  "This_Week" : 20,
  "Last_Week" : 20,
  "End_2019" : 19,
  "Name" : "Shane Lowry",
  "Average_Points" : 4.2695,
  "Total_Points" : 213.48,
  "Events_Played_Divisor" : 50,
  "Points_Lost_2020" : -37.09,
  "Points_Gained_2020" : 27.69,
  "Events_Played_Actual" : 50,
  "Pro/Am" : "Pro"
}
```

Fig. Colección sobre datos de Kaggle

A continuación, se muestra las colecciones creadas en la base de datos “pythondb”.

Collections						
CREATE COLLECTION						
Collection Name	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
facebook	84	916.6 B	75.2 KB	1	20.0 KB	
firstcollection	4	328.5 B	1.3 KB	1	20.0 KB	
kaggle	9,000	260.0 B	2.2 MB	1	100.0 KB	
tiktok	270	3.2 KB	856.2 KB	1	20.0 KB	
twitter	360	7.1 KB	2.5 MB	1	24.0 KB	

Fig. Colecciones de la base de datos “pythondb”

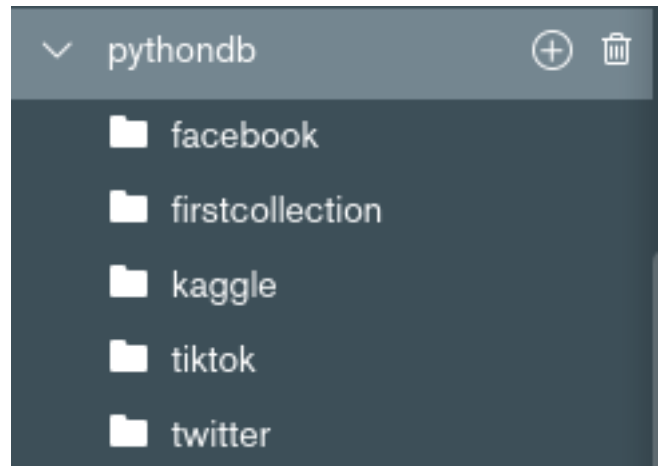


Fig. BD\_Python

Los datos han sido recopilados en su mayoría por scripts en Python, aunque también se tienen documentos descargados de la base de datos del CNE



Fig. Descarga de datos

```
import pandas as pd
import mysql.connector

bdprev = pd.read_excel('candidatos.xlsx')

mydb = mysql.connector.connect(user='root', password='Casv1752.', host='localhost', database='politica')

Cargo=bdprev['DIGNIDAD_NOMBRE']
Provincia=bdprev['CANDIDATO_PROVINCIA_NOMBRE']
Nombre=bdprev['CANDIDATO_NOMBRE']
Partido=bdprev['CANDIDATO_NOMBRE_RESULTADOS']
Gana=bdprev['CANDIDATO_ESTADO']

mycursor = mydb.cursor()
mycursor.execute("CREATE TABLE Candidatos (id INT AUTO_INCREMENT PRIMARY KEY, NOMBRE_CANDIDATO VARCHAR(255), PROVINCIA VARCHAR(255), PARTIDO VARCHAR(255), CARGO VARCHAR(255), VENCEDOR VARCHAR(255))")
sql = "INSERT INTO Candidatos (NOMBRE_CANDIDATO, PROVINCIA, PARTIDO, CARGO, VENCEDOR) VALUES (%s, %s, %s, %s, %s)"
val=[]
for i in range(Cargo.len()):
    val.append((Nombre[i], Provincia[i], Partido[i], Provincia[i], Cargo[i], Gana[i]))
mycursor.execute(sql, val)
mydb.commit()
print(mycursor.rowcount, "registro insertado")
mydb.close()
```

Fig. Script MySql Candidatos



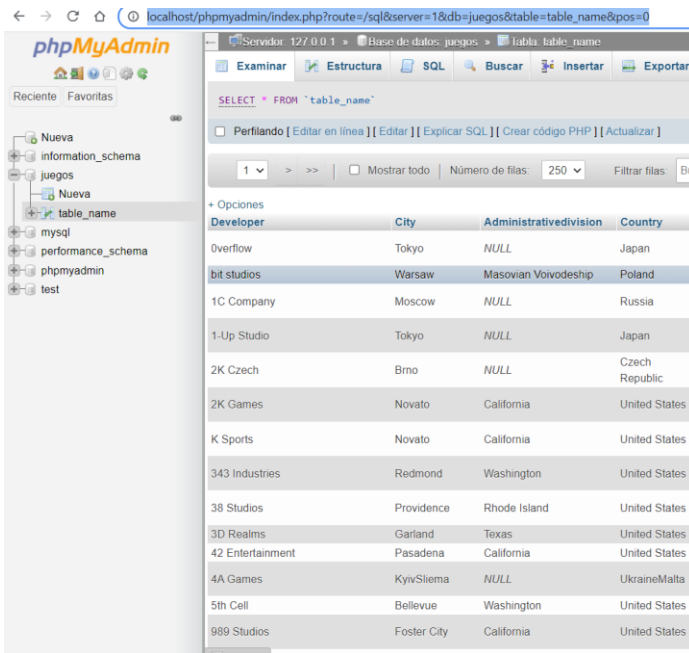
Reemplazar esta línea con su número de documento de identificación (hacer doble clic aquí para editar) < 6

```
1
2 import pandas as pd
3 import mysql.connector
4
5 mydb = mysql.connector.connect(user="root",password="Casvi752.",host="localhost",database="politico")
6
7 bdprev = pd.read_excel('organizaciones_pol.Xlsx')
8
9 Provincia=bdprev['OP_PROVINCIA_NOMBRE']
10 Ambito=bdprev['OP_AMBITO']
11 Partido=bdprev['OP_PARTIDO']
12
13 mycursor = mydb.cursor()
14
15 mycursor.execute("CREATE TABLE Partidos (id INT AUTO_INCREMENT PRIMARY KEY, PARTIDO VARCHAR(255),
16 sql = "INSERT INTO Partidos (PARTIDO, PROVINCIA, AMBITO ) VALUES (%s, %s, %s)"
17 val=[]
18 for ite in range (Provincia.len()):
19     val.append((Partido[ite], Provincia[ite], Ambito[ite]))
20 mycursor.execute(sql, val)
21 mydb.commit()
22 print(mycursor.rowcount, "registro insertado")
23 mydb.close()
```

Fig. Script MySQL Partidos.

## VII. VISUALIZACIÓN DE LA INFORMACIÓN

Para la visualización de la información, se necesitó almacenar los datos en Elasticsearch. A continuación, se presenta el proceso realizado con la base de datos “juegos” en MySQL



The screenshot shows the phpMyAdmin interface. On the left, the database 'juegos' is selected. The main area displays the 'table\_name' table structure with columns: Developer, City, Administrative division, and Country. Below the structure, a table of data is shown with 10 rows and 4 columns.

Developer	City	Administrative division	Country
Overflow	Tokyo	NULL	Japan
bit studios	Warsaw	Masovian Voivodeship	Poland
1C Company	Moscow	NULL	Russia
1-Up Studio	Tokyo	NULL	Japan
2K Czech	Brno	NULL	Czech Republic
2K Games	Novato	California	United States
K Sports	Novato	California	United States
343 Industries	Redmond	Washington	United States
38 Studios	Providence	Rhode Island	United States
3D Realms	Garland	Texas	United States
42 Entertainment	Pasadena	California	United States
4A Games	Kyiv/Sliema	NULL	Ukraine/Malta
5th Cell	Bellevue	Washington	United States
989 Studios	Foster City	California	United States

Fig. BD juegos en MySQL

Cada base de datos fue indexada a Elasticsearch mediante logstash, para lo cual fue necesario de los archivos de configuración. Con el cliente Cerebro se pudo obtener una GUI para una visualización de Elasticsearch.

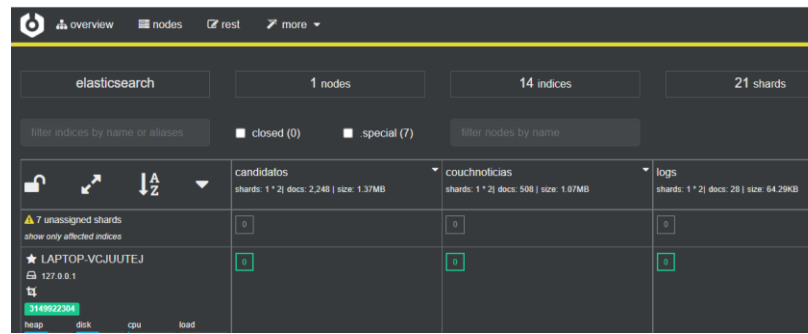
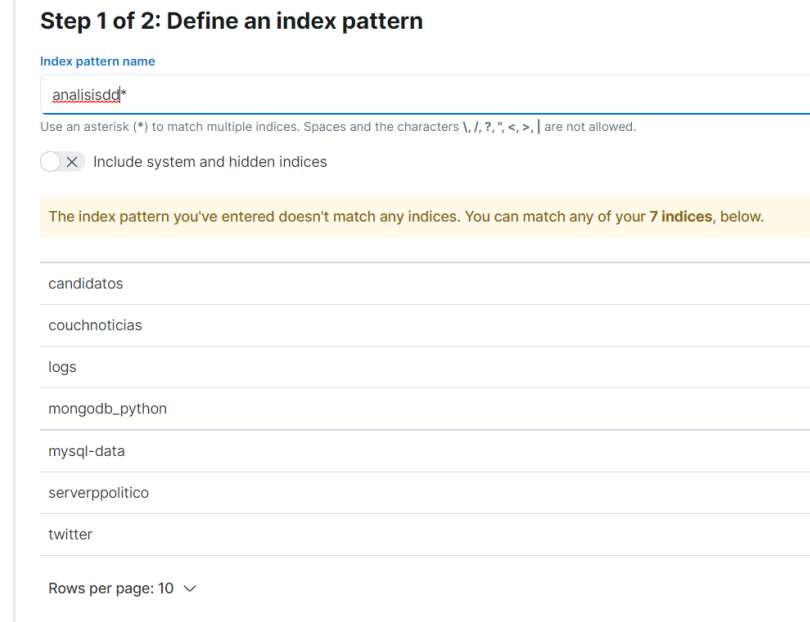


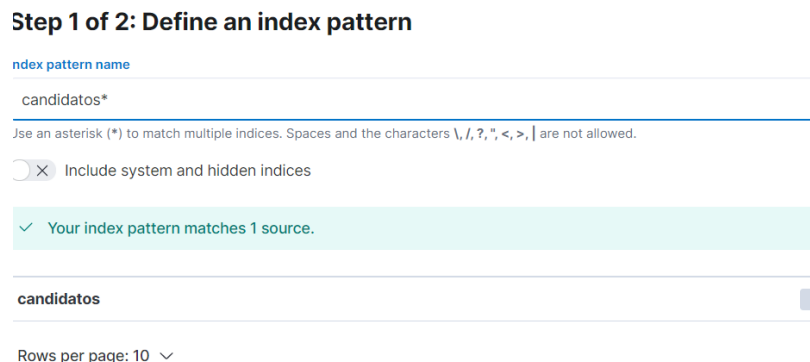
Fig. Índices en Elasticsearch.

Seguidamente, se continuó con la visualización y análisis estadístico mediante Kibana.



The screenshot shows the Kibana 'Step 1 of 2: Define an index pattern' form. The 'Index pattern name' field contains 'analisisdd\*'. Below the field, there is a message: 'The index pattern you've entered doesn't match any indices. You can match any of your 7 indices, below.' A list of indices is shown: 'candidatos', 'couchnoticias', 'logs', 'mongodb\_python', 'mysql-data', 'serverppolitico', and 'twitter'. The 'Rows per page' is set to 10.

Fig. Definición de un índice pattern.



The screenshot shows the Kibana 'Step 1 of 2: Define an index pattern' form. The 'Index pattern name' field contains 'candidatos\*'. Below the field, there is a message: 'Your index pattern matches 1 source.' A list of sources is shown: 'candidatos'. The 'Rows per page' is set to 10.

Fig. Creación de un índice pattern.

Se eligió la opción de análisis en tiempo real, para obtener una comparación y una visualización del cambio de los tweets.

Reemplazar esta línea con su número de documento de identificación (hacer doble clic aquí para editar) < 7

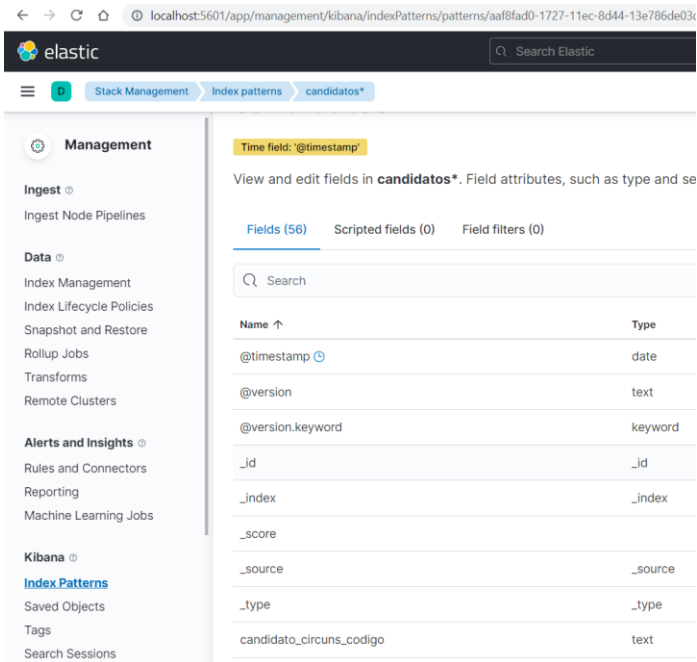


Fig. Creación de un index pattern.

Una vez creado el index pattern, se creó las visualizaciones para obtener un mejor análisis estadístico mediante los gráficos que proporciona Kibana.

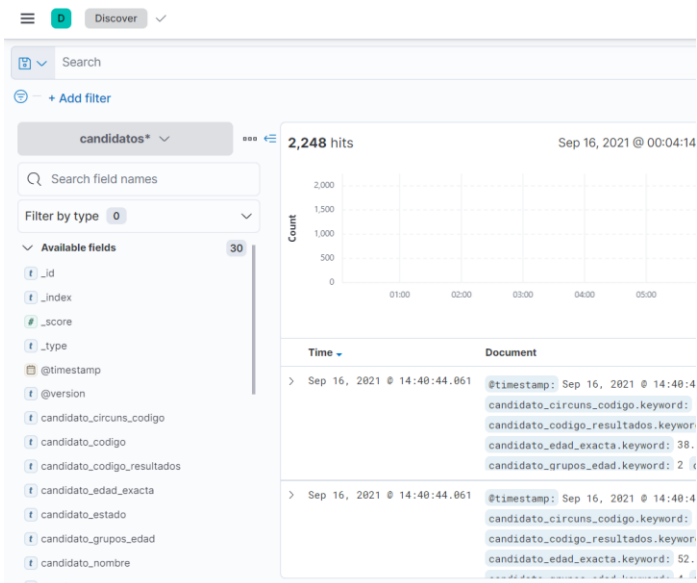


Fig. Gráfico en tiempo real

Finalmente, se muestra el gráfico de área del índice de Elasticsearch “juegos” obtenido mediante Kibana.

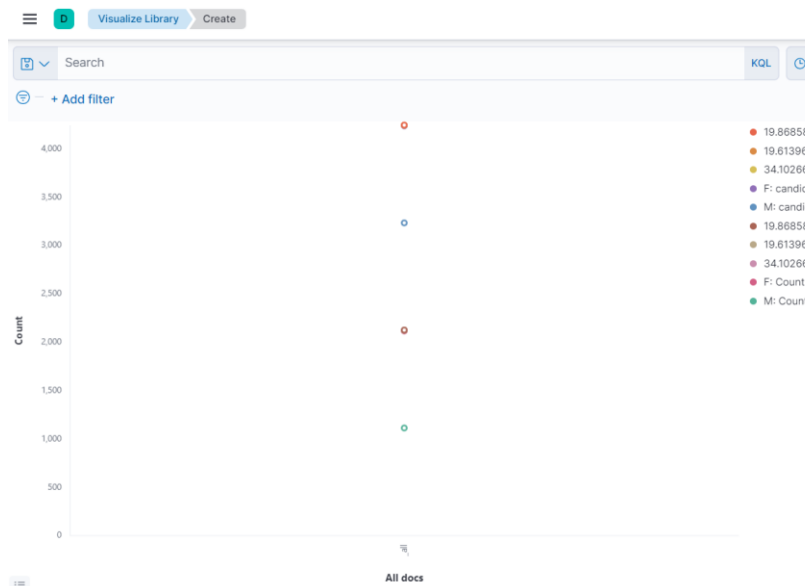


Fig. Gráfico de área

## VIII. ANÁLISIS DE LA INFORMACIÓN

### Juegos

La tabla de juegos nos muestra una tendencia a ser desarrollados en los Estados Unidos, seguido por el reino Unido. Esto contraste con la creencia popular de que son desarrollados más en Asia.

### Python

Se realizo con una recolección de datos en tiempo real en Ecuador. Por lo tanto, los datos obtenidos nos muestran una tendencia en los Quiteños

### Noticias

Según los datos recolectados la mayoría de las noticias publicadas no especifican su ubicación, luego le prosigue Río de Janeiro y Estados Unidos.

### Pulso Político

### (Ciudades)

Reemplazar esta línea con su número de documento de identificación (hacer doble clic aquí para editar) < 8

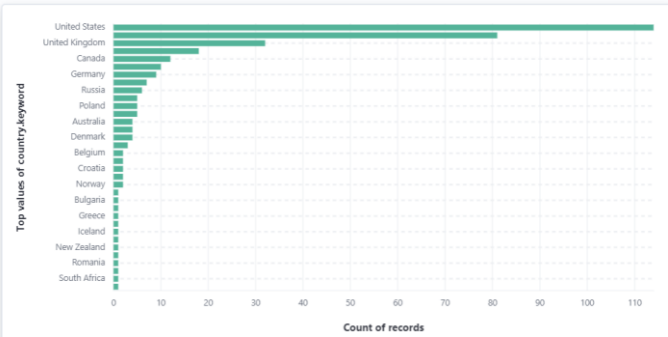
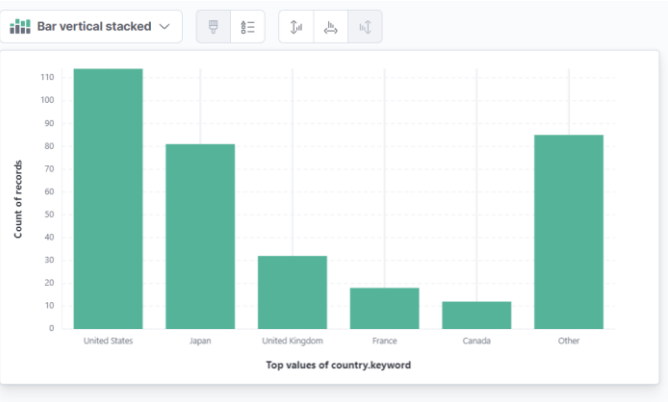
De la muestra de las ciudades se encontró una gran cantidad de partidos y movimientos que participan en estas localidades para un puesto en la asamblea, recalcando que muchos contaban con un gran número de candidatos.

(Provincias)

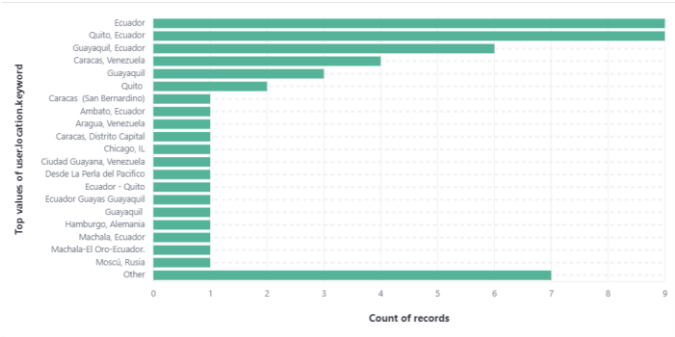
Aumentando la muestra se concluye que en las provincias del Guayas y Pichincha la participacion es sumamente mayor, demostrando el interes de los movimientos en tener influencia en estos sectores. Caso contrario sucede en el oriente con provincias como Napo y Morona Santiago cuya poblacion es menor pero poseen mas recursos que otras zonas del ecuador.

IX. RESULTADOS OBTENIDOS

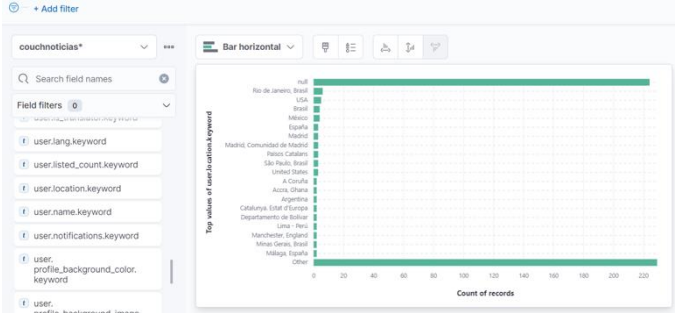
Juegos



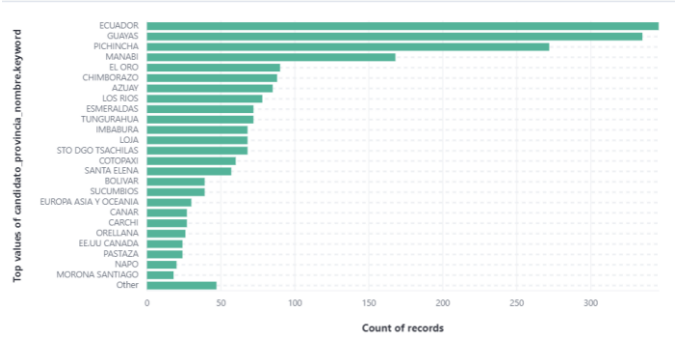
python



Noticias



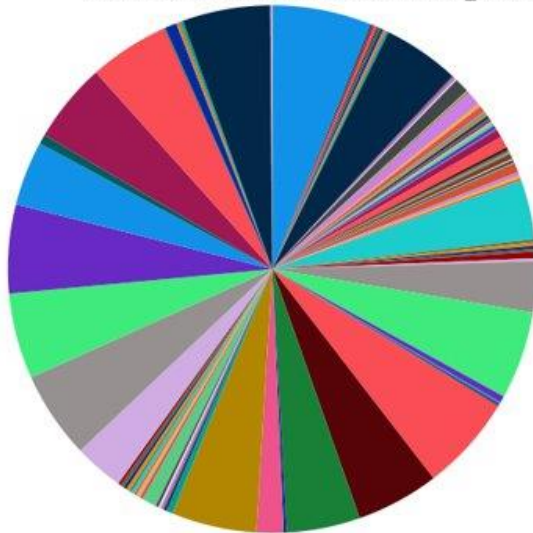
candidatos



PULSO POLITICO



Gráfico circular Recuento de CANDIDATO\_NOMBRE\_RESUL

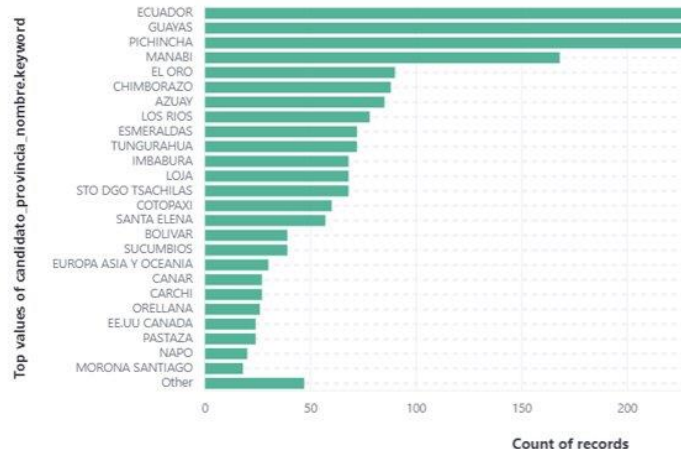


- Recomendaciones
  - Como experiencia del proyecto, se recomienda utilizar Elasticsearch en la nube; pues generar los procesos de logstash, kibana y cerebro exigen recursos a la máquina.
  - Se recomienda que, no optar a Elasticsearch como alternativa a una base de datos relacional.

## XI. DESAFIOS Y PROBLEMAS ENCONTRADOS

- Falta de conocimiento en cuanto a DataLake.
- Empleo de API's.
- Herramientas necesarias para ELK Encontrar métricas adecuadas.

## PULSO POLITICO POR PROVINCIAS



## REFERENCES

- [1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
- [2] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
- [3] H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.
- [4] B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
- [5] E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.
- [6] J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.
- [7] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private
- [8] 876—880. Available: <http://www.halcyon.com/pub/journals/21ps03-vidmar>

## X. CONCLUSIONES Y RECOMENDACIONES

- Conclusiones
  - La búsquedas de data Con Elasticsearch se obtuvo un manejo de los modelos de los datasets.
  - Mediante este análisis de datos se puede obtener diferentes interpretaciones dependiendo de las métricas que se empleen.
  - En cuanto a resultados obtenidos, podemos mencionar que en el caso de la política aun habiendo pasado las elecciones las personas siguen hablando de ellos, ya sea a favor o en contra.