

# Informe proyecto final

Casa Velasco Edwin David  
Quishpe Catagña Jonathan Andrés  
Herrera Gómez Ricardo Andrés  
Análisis de Sistemas Informáticos  
Escuela Politécnica Nacional  
edwin.casa01@epn.edu.ec  
jonathan.quishpe01@epn.edu.ec  
ricardo.herrera01@epn.edu.ec

## I. INTRODUCCIÓN

En el presente documento se describen los pasos realizados para la recopilación, disposición e interpretación de datos, con lo cual se procederá a la creación de un Data warehouse que os brinde la posibilidad de diseñar un acceso para explicar la información obtenida de manera clara y sencilla para un mejor entendimiento, mediante el uso de graficas de diferentes tipos

## II. DESARROLLO

### 1. Definición del caso de estudio.

Implementar una arquitectura Data Warehouse, mediante la cual se analice diferente información que será recolecta de tweets con distintos parámetros, para así satisfacer todas las temáticas propuestas. El análisis de la información se la realizará con las diferentes herramientas utilizadas en clase como son: Elasticsearch, CouchDB, Kibana, logstash, para así poder realizar representaciones graficas con distintos parámetros. La información por analizar será actual, para así poder saber el comportamiento de las personas dependiendo las temáticas, para saber en qué momento postean algo en redes sociales, el uso que dan a las redes sociales en este caso Twitter, para expresar sus emociones.

### 2. Objetivo General y específicos.

#### Objetivo General.

Implementar una arquitectura Data Warehouse, para la recolección y análisis de datos con herramientas usadas en clase, además de la generación de gráficos para visualizar la información de forma más clara.

#### Objetivos Específicos.

- Recopilar datos en CouchDB mediante la API de Twitter.
- Diseñar la arquitectura Data Warehouse a utilizar.
- Transformación de los datos
- Creación de índices mapeados
- Generar un Dashboard con las diferentes visualizaciones

### 3. Descripción del equipo de trabajo.

El equipo de trabajo se encuentra conformado por:

- Casa Velasco Edwin David
- Quishpe Catagña Jonathan Andrés
- Herrera Gómez Ricardo Andrés

Estudiantes de la materia Base de Datos multidimensionales, poseen conocimientos en el uso de

herramientas de análisis de datos como son couchDB, ElasticSearch, Kibana, entre otras.

#### Actividades realizadas por integrante.

En la Ilustración 1, se puede observar las actividades realizadas por cada uno de los integrantes.

INTEGRANTES	ACTIVIDADES					
	RECOLECCIÓN DE DATOS	TRANSFORMACIÓN DE DATOS	CREACIÓN VISUALIZACIONES	CREACIÓN DASHBOARD	MAPPING E INDICES	INFORME
Jonathan Quishpe						
Ricardo Herrera						
David Casa						

Ilustración 1 Actividades por integrante

#### 4. Cronograma de actividades

Las diferentes actividades por realizarse en el proceso de desarrollo del proyecto se pueden observar en la Ilustración 2.

ACTIVIDADES	Inicio	Fin	SEMANAS			
			1	2	3	4
RECOLECCION DE DATOS						
TRANSFORMACION DE DATOS						
CREACION VISUALIAZACIONES						
CREACION DASHBOARD						
INFORME						

Ilustración 2 Cronograma de actividades

#### 5. Recursos y herramientas utilizadas



Ilustración 3 Herramientas

Los recursos y herramientas utilizados para la realización del proyecto final de la materia base de Datos Multidimensionales fueron:

- **CouchDB 2.1.2**, Apache CouchDB, comúnmente llamada 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".<sup>1</sup> Se trata de una base de datos NoSQL que emplea JSON para almacenar los

datos, JavaScript como lenguaje de consulta por medio de MapReduce y HTTP como API.

- **Api Twitter**, la cual ayuda a los desarrolladores a la obtención de datos de manera sencilla, mediante una conexión con la app tweet, e instalando los paquetes necesarios para su correcta ejecución
- **ElasticSearch**, es un servidor de búsqueda basado en Lucene. Provee un motor de búsqueda de texto completo, distribuido y con capacidad de multi-tenencia 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 ideal para la agregación de logs, visualización y análisis, aunque no está diseñado específicamente para este propósito.
- **Logstash**, es una herramienta para la administración de logs. Esta herramienta se puede utilizar para recolectar, parsear y guardar los logs para futuras búsquedas.
- **Computadores e internet**: usados para la realización de todo el proyecto, son fundamentales debido a que sin un computador no se puede realizar los objetivos, además de la conexión a internet para así poder capturar datos.

## 6. Arquitectura de la solución

La arquitectura construida para el análisis de datos recolectados se puede observar en la Ilustración 4.

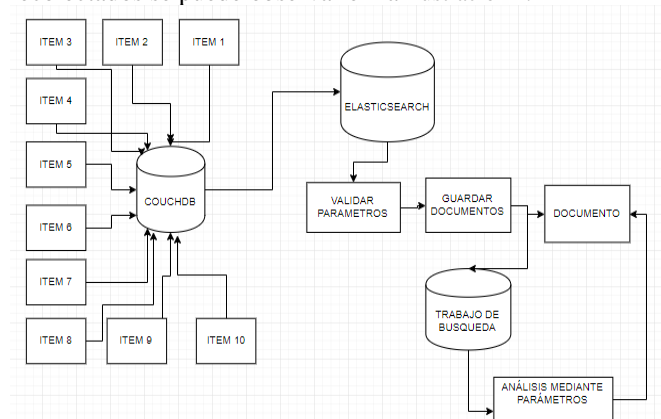


Ilustración 4 Arquitectura construida

## 7. Extracción de datos

A continuación, se presenta la cantidad de datos recolectados para analizar cada situación propuesta:

Tráfico vehicular en las 5 principales ciudades del Ecuador:

Eventos deportivos en los principales estadios de Ecuador:

Pulso político en 5 ciudades de Ecuador:

Top 10 twitteros en 5 ciudades de Ecuador:

Top 10 quejas en el Ecuador:

Actividades y hobbies:

Conciertos y eventos públicos:

Tema definido por el estudiante:

Restaurantes y sitios de esparcimiento:

Eventos o noticias mundiales:

## 8. Análisis de la información

### Tráfico vehicular en las 5 principales ciudades del Ecuador.

En base a la información recolectada y haciendo uso de un gráfico de puntos de localización, se puede observar que la personas utilizan los medios sociales para emitir sus molestias en cuanto en referencia al tráfico vehicular, cabe recalcar que la información obtenida no solamente se refiere a tráfico vehicular, ya que al momento de la recolección de datos se filtró por la palabra tráfico en la diferentes ciudad lo que ocasiona que la información no se refiera a un solo tipo de tráfico, como por ejemplo en la Ilustración 5, se puede observar que se refiere a un tráfico diferente.

### Eventos deportivos en los principales estadios de Ecuador

La información recolectada en cuanto a eventos deportivos se la represento mediante un gráfico de forma pastel con la palabra deportes, en la cual se puede observar que las personas postean sobre diferentes eventos que se realizan en los estadios del Ecuador, Ilustración 6, para un análisis más a profundidad se debería elegir un tipo de evento puntual, ya que los tweets no proporcionan información muy puntual.

### Pulso político en 5 ciudades de Ecuador

En referencia a los datos obtenidos acerca de política o pulso político en 5 ciudades de quito, las personas debaten o comentan sobre política en las ciudades de Quito, Guayaquil, Cuenca, Quevedo, esta información es extraída mediante un gráfico de geolocalización la cual se la puede observar en la Ilustración 7.

### Top 10 twitteros

En cuanto a la información sobre el topten de twitteros en ecuador, se puede observar en la Ilustración 8, en la parte derecha las personas mas mencionadas en los tweets, o las personas las cuales más postean, entre esa lista se puede observar a el ex presidente del Ecuador, así como a personas involucradas en el mundo de la actuación o televisión.

### Top 10 quejas

En cuanto a la búsqueda de un topten de quejas existe gran cantidad de estas, y para la elección de las más posteadas se debe realizar un análisis más a profundidad, o realizar una búsqueda más concreta en referencia a información obtenida en la red sobre cuáles son las quejas más emitida en las diferentes redes sociales, observar resultados en la Ilustración 9.

### Actividades y hobbies

Las actividades y hobbies que las personas realizan a menudo las personas son posteadas, en la Ilustración 10, se puede observar que la cantidad de posts sobre actividades y hobbies es alta, se puede observar que argentina es la ciudad que más ha posteoado sobre dichos temas.

### Conciertos y eventos públicos

En cuanto a concierto y eventos públicos existe una gran cantidad de posts, en la ciudad Quito ya que en estas fechas se vive el VAQ, o verano de las artes por lo que la cantidad de tweets representan o pertenecen a este tema, se puede observar la gráfica en la Ilustración 11.

### Tema definido por el estudiante

Los temas definidos fueron PUBG y Fornite, los cuales son juegos en línea que en la actualidad son el boom en cuanto a

juegos de supervivencia, la información obtenida en diferentes países es elevada, en el Ilustración 12 se puede observar los diferentes países que postean sobre dichos juegos.

### Restaurantes y sitios de esparcimiento

Mediante un gráfico de tipo HeadMap, se puede observar que los lugares en los que más se postea sobre este tópico es en las ciudades de Guayaquil y Quito como se puede observar en la Ilustración 13.

### Eventos o noticias mundiales

En cuanto a este tópico, existen una gran cantidad de tweets, los cuales se pueden observar en la Ilustración 14, noticias en todo idioma se puede recolectar en la red, en un tiempo muy corto.

## 9. Visualización de la información

A continuación, se presentan las gráficas generadas a partir de los datos recolectados, cada grafico representa un ítem del literal 7.

### Tráfico vehicular

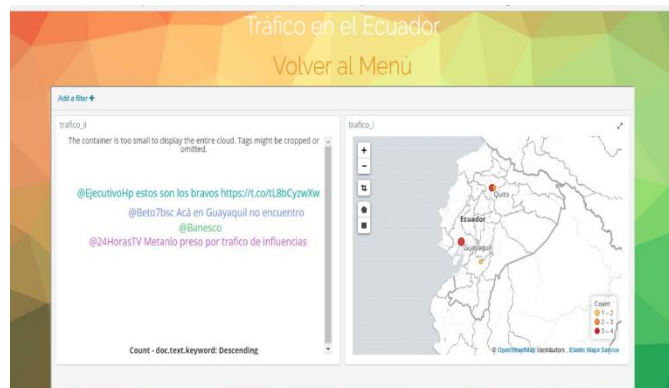


Ilustración 5 Tráfico vehicular

### Eventos deportivos

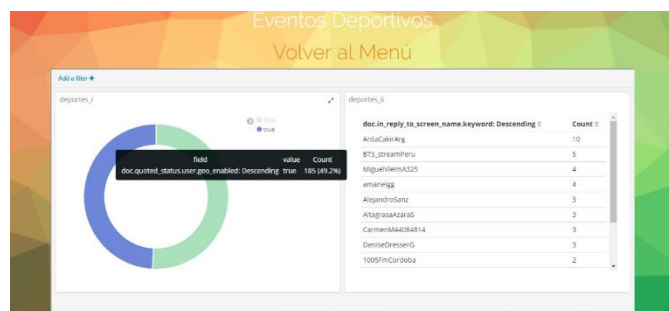


Ilustración 6Eventos deportivos

### Pulso político



Ilustración 7Pulso político

### Top 10 twittereros



Ilustración 8Top 10 twittereros

### Top 10 quejas



Ilustración 9Top 10 quejas

### Actividades y hobbies

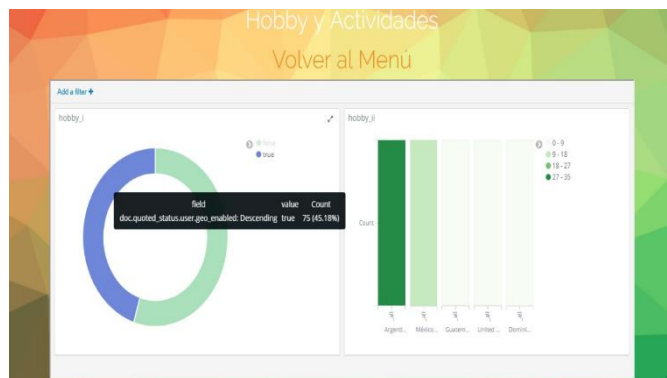


Ilustración 10Actividades y hobbies

### Conciertos y eventos públicos

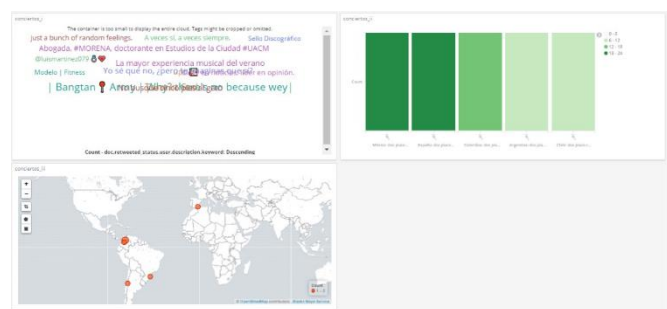


Ilustración 11Conciertos y eventos públicos

## Tema definido por el estudiante

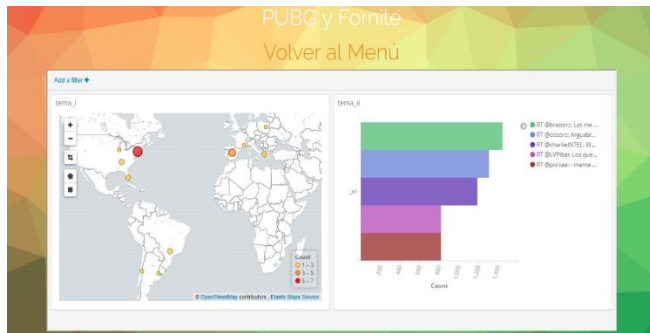


Ilustración 12 PUBG Y Fortnite

## IV.DESAFÍOS Y PROBLEMAS ENCONTRADOS

Al momento del desarrollo del proyecto, se tuvo problemas ya que en un inicio la recolección de datos se la realizó en la última versión de couchdb, la cual no permite o no se pudo enlazar las bases de datos removiendo de carpetas, la solución se la está analizando ya que los datos recolectados la última versión del gestor de base de datos es una cantidad considerable.

## Restaurantes y sitios de esparcimiento

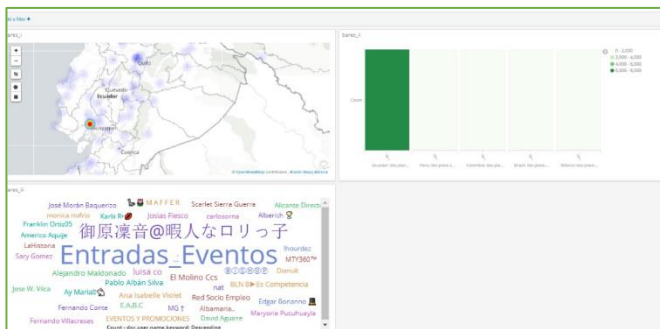


Ilustración 13 Restaurantes y sitios de esparcimiento

## Eventos o noticias mundiales



Ilustración 14 Eventos o noticias mundiales

## III.CONCLUSIONES Y RECOMENDACIONES

El uso de una base de datos multidimensional es de gran ayuda para el análisis de datos, ya que se pueden analizar los datos desde diferentes puntos de vistas, para así poder tomar decisiones más exactas, basándose en información histórica arrojada por el estudio.

Al finalizar este proyecto se dimensiona la importancia del concepto de Base de datos multidimensionales en la actualidad, la oportunidad que nos entregan al momento de analizar datos a través del tiempo o responder consultas que con una base de datos relacional ocuparía demasiados recursos y llevaría mucho tiempo tomando en cuenta la gran cantidad de información y las operaciones que se tendrían que realizar para obtener estos resultados.