# Proyecto Big Data Processing

El desarrollo de este proyecto final de la sesión de Big Data Processing contempla el uso de Spark para el procesamiento de datos, haciendo uso de una infraestructura que contempla Apache Kafka, PostgreSQL y almacenamiento de datos en formato parquet.

Se hace uso de batch processing, de streaming, agregaciones, joins, a modo de cumplir con los requerimientos del proyecto.

## Parte 1 Speed layer.

Capa de procesamiento de streaming, toma la información en tiempo real y la preprocesa.

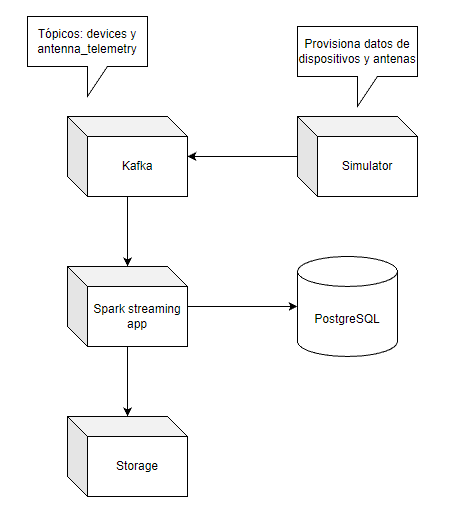
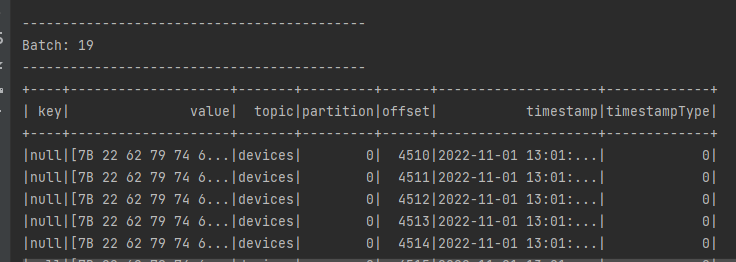


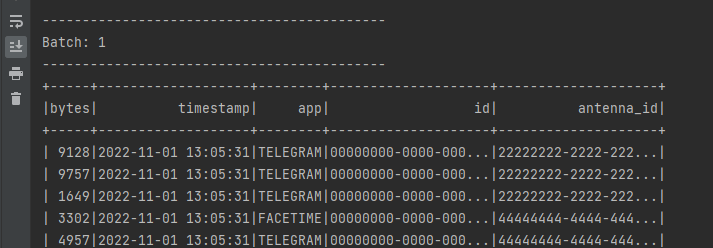
Figura Arquitectura de speed layer

Bloques de la speed layer

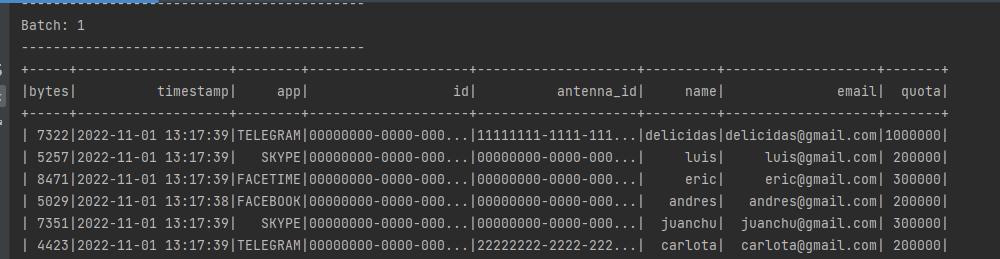
1. Consumo de Kafka



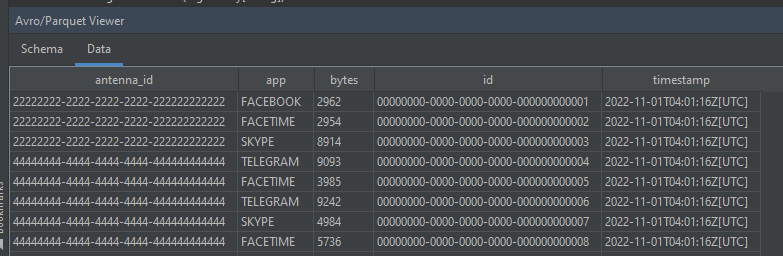
1. La información se transforma usando un schema json



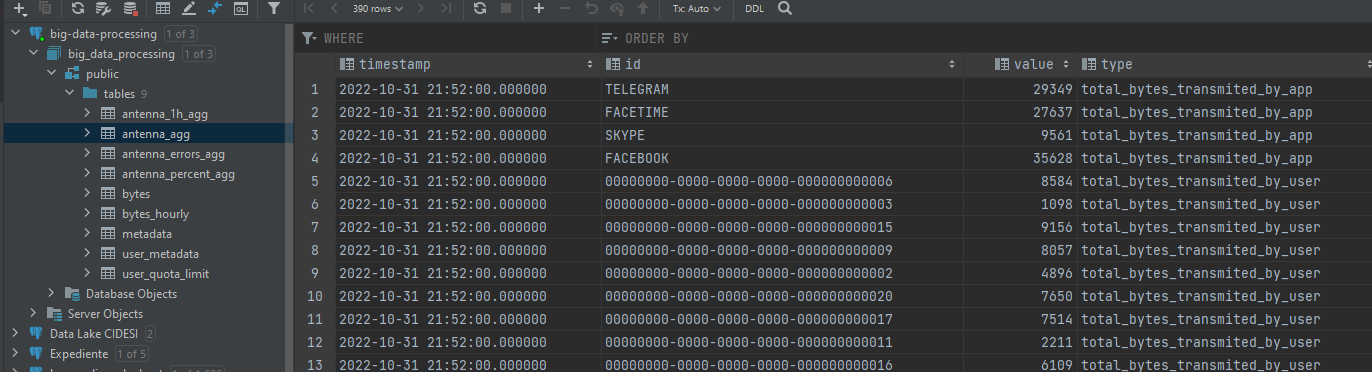
1. Se enriquece el dataframe usando los metadatos adionales



1. La información se almacena en formato parquet



1. La información se va a PostgreSQL



## Parte 2 Batch layer

Capa de procesamiento por lotes

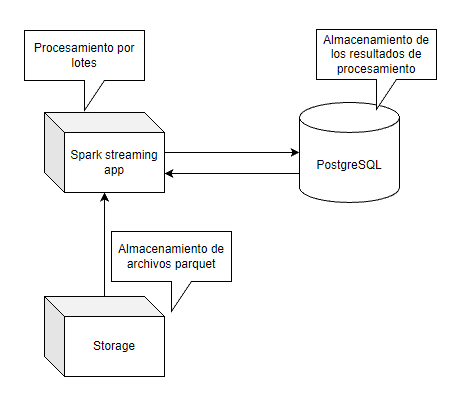
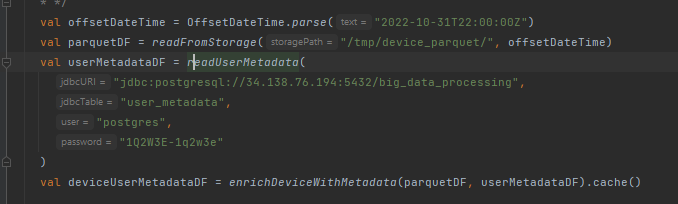


Figura Flujo de procesamiento de archivos parquet

Bloques de procesamiento de información

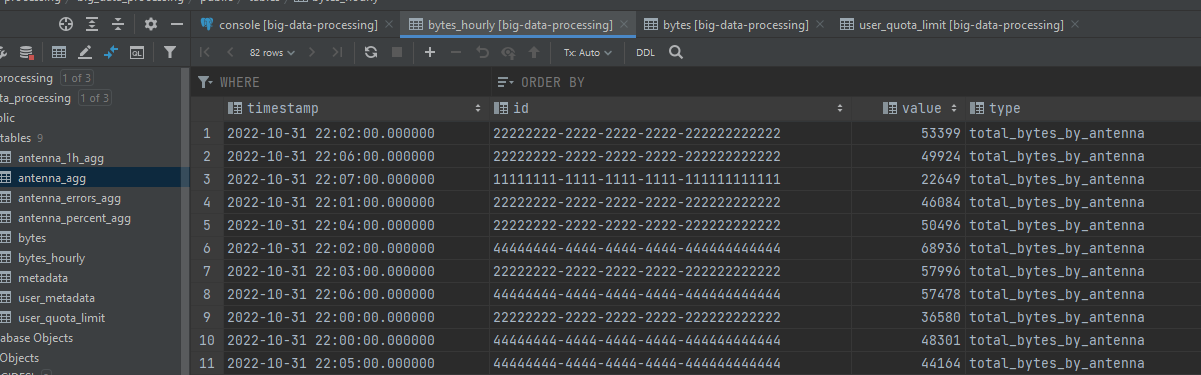
La información se trata de la misma forma en la mayoría de los casos, sin embargo, en el cumplimiento de los requerimientos, la información se toma del almacenamiento, en formato parquet, se adicionan los metadatos como sigue a continuación:



Después de que se enriquecen los datos, se realiza el computo de los requerimientos, para así mandarlos a la base de datos



La información en base de datos ya está computada y se clasifica por tipo de información



Para el caso de computo del limite de cuota se va a una tabla diferente

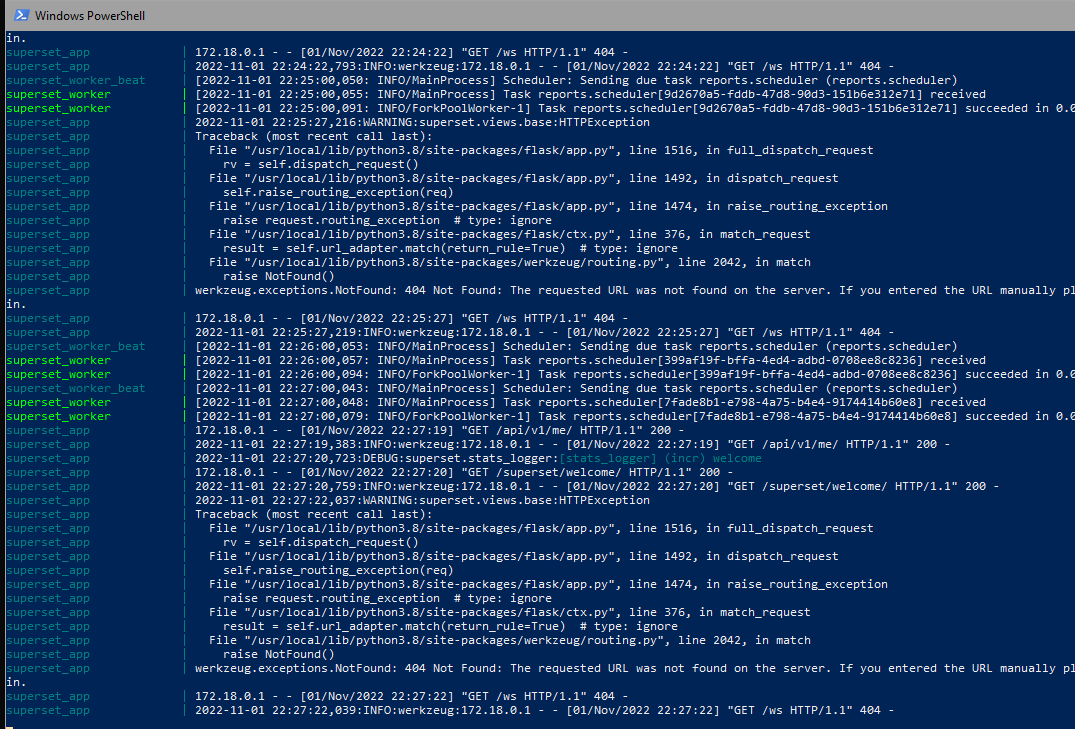
Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

## Parte 3 Serving layer

Uso de apache superset

Se intentó hacer uso de apache superset, aunque se tuvieron algunos inconvenientes



Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Se probo en Windows, se va a probar en instancias de GCP y en otros ordenadores