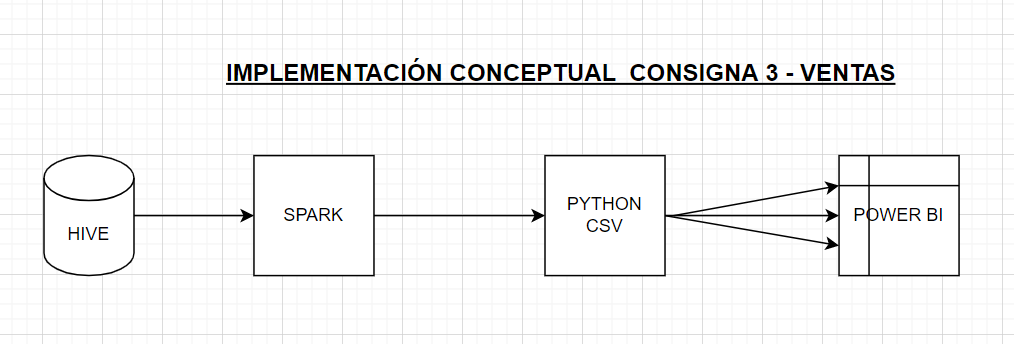
**DOCUMENTACIÓN CONSIGNA NRO 3 – VENTAS**

**1. Planificación**

* Definición de alcance
  + Consigna Nro3 - Ventas
  + Preguntas de negocio:
    - 1. FACTURACIÓN TOTAL  POR FAMILIA DE PRODUCTO
      2. TOP 10 DE LOS PRODUCTOS MÁS VENDIDOS POR SUCURSAL
      3. QUÉ SUCURSAL VENDIÓ MÁS POR CANTIDAD
* Descripción de la solución
  + Diagrama de arquitectura



* + Flujo de datos
    - Herramienta usada en cada etapa

El proyecto consiste en realizar el procesamiento de datos con las herramientas propuestas en los cursos, con los cuales se debe desarrollar una flujo de datos que procesen y respondan a  las distintas consignas, las cuales son:

* FACTURACIÓN TOTAL  POR FAMILIA DE PRODUCTO
* TOP 10 DE LOS PRODUCTOS MÁS VENDIDOS POR SUCURSAL-
* QUÉ SUCURSAL VENDIÓ MÁS POR CANTIDAD

Como equipo optamos por responder 3 preguntas y a continuación describiremos cual es la lógica del flujo propuesto por nosotros.

El origen de los datos serán archivos.CSV y se utilizarán las siguientes herramientas para su carga, implementación de lógica y finalmente su visualización mediante reportes:

**HIVE:** herramienta utilizada para la carga de los archivos en tablas stg.  
**Spark:** herramienta con la cual se implementará la lógica de negocio para la generación de los tablones de información.  
**Python:** herramienta con la que se generaran archivos CSV con el fin de utilizarlos como input de Power BI.  
**Power BI:** herramienta de visualización con la cual se desplegará la información que responderá a la consigna de venta.

* + - Datos de entrada y salida (archivos, tablas, tópicos, etc.)

**DATOS DE ENTRADA:**

empleados.csv

fact.csv

sucursales.csv

producto.csv

**DATOS DE SALIDA:**

academia\_lnd.l0507199\_empleados

academia\_lnd.l0507199\_fact

academia\_lnd.l0507199\_productos

academia\_lnd.l0507199\_sucursales

**ARCHIVOS.CSV:**

Tablon1.csv

Tablon2.csv

**POWER BI:**

Reporte.facturación\_total\_por\_familia

Reporte.top\_10\_productos\_vendidos\_sucursal

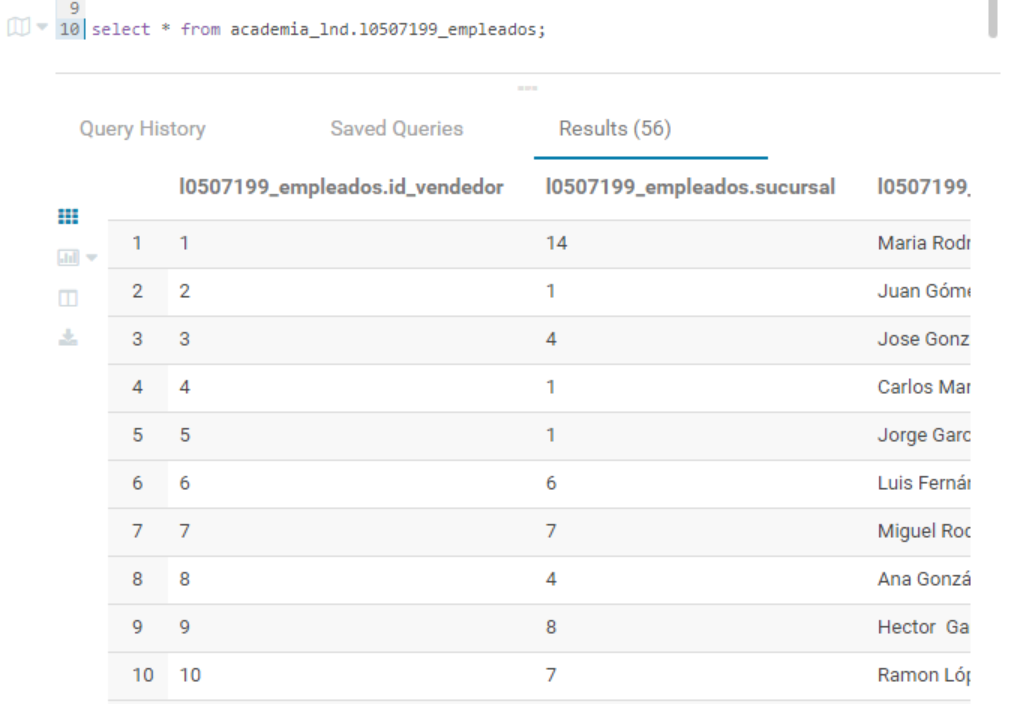
Reporte.ranking\_ventas\_sucursal

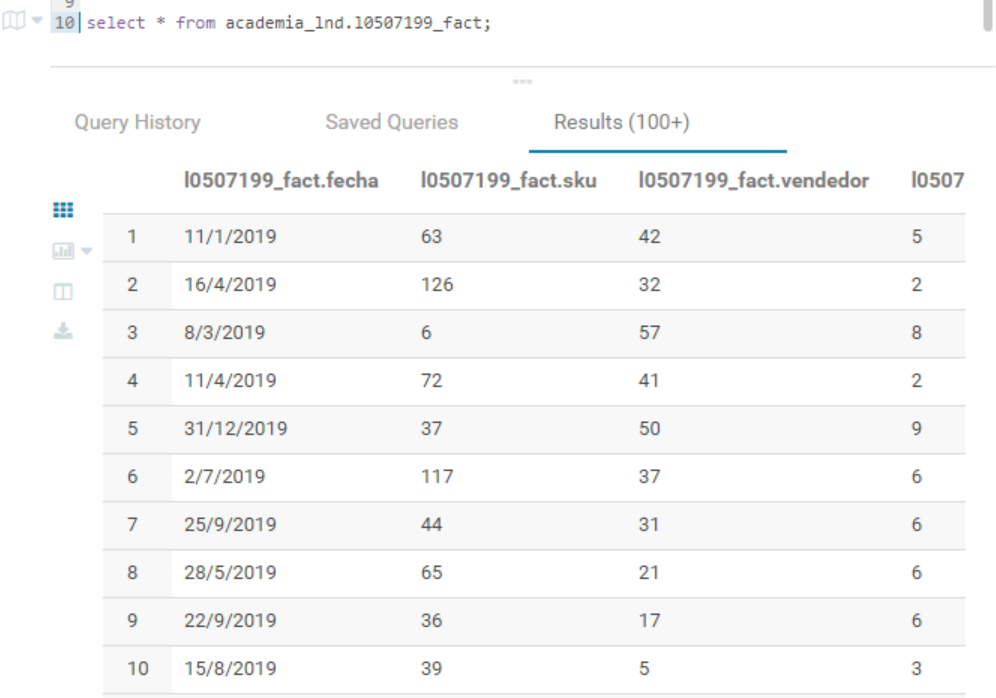
* + - Procesamiento

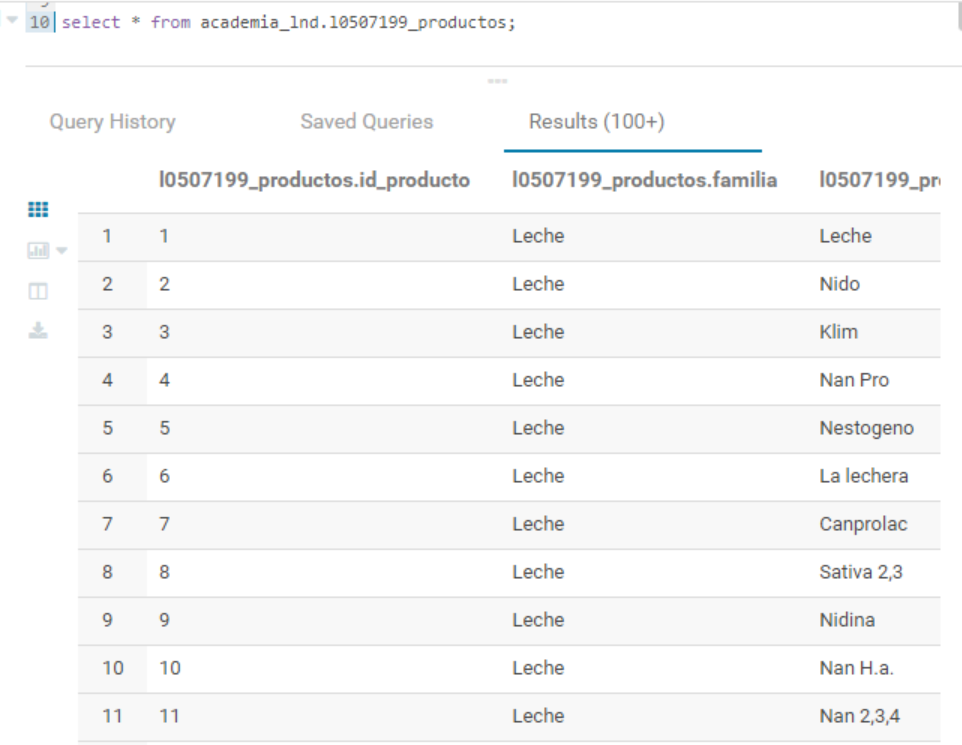
Se consumirán los 4 archivos.CSV (empleados, fact, producto, sucursales) los cuales serán cargados mediante HIVE en tablas staging con la finalidad de poder realizar la implementación de lógica con Spark, así generando 2 tablones (Tablon1.csv, Tablon2.csv) que responderán a nuestra consigna relacionada con ventas. Con Python se traspasará la data de los tablones a los archivos.CSV para ser tomados finalmente con Power BI y así generar la reportería correspondiente de nuestro proyecto.

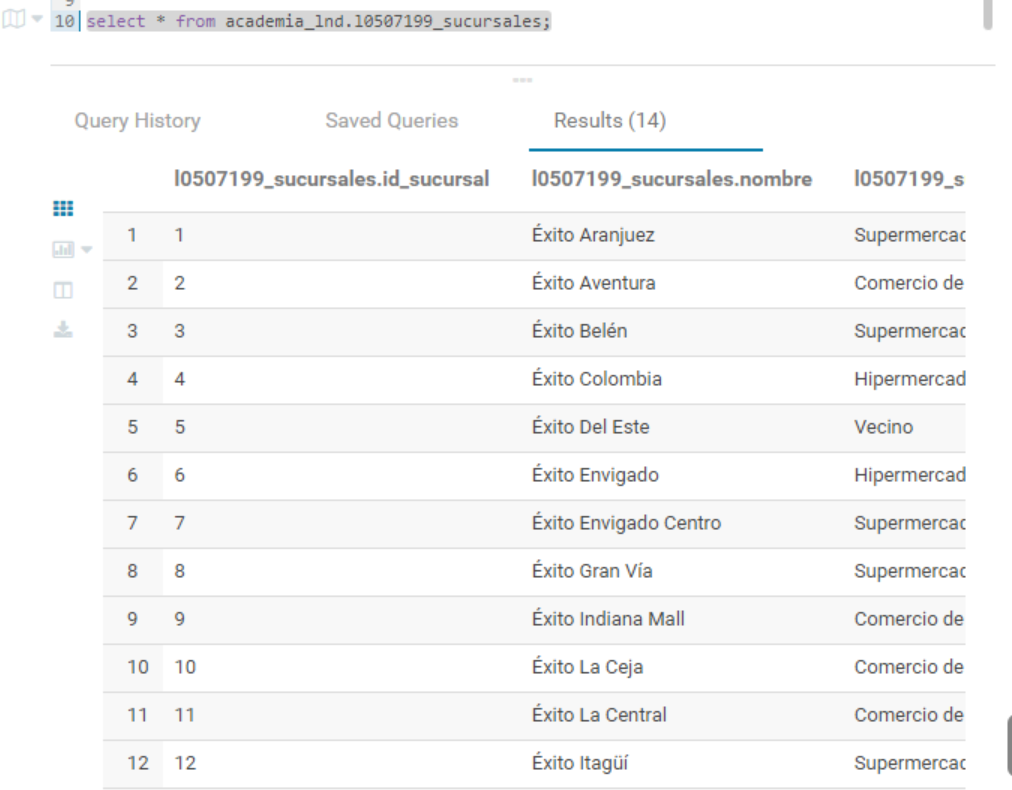
**2. Implementación**

* 07/08/2022 se trabajó en la creación y carga de las tablas en Hive.

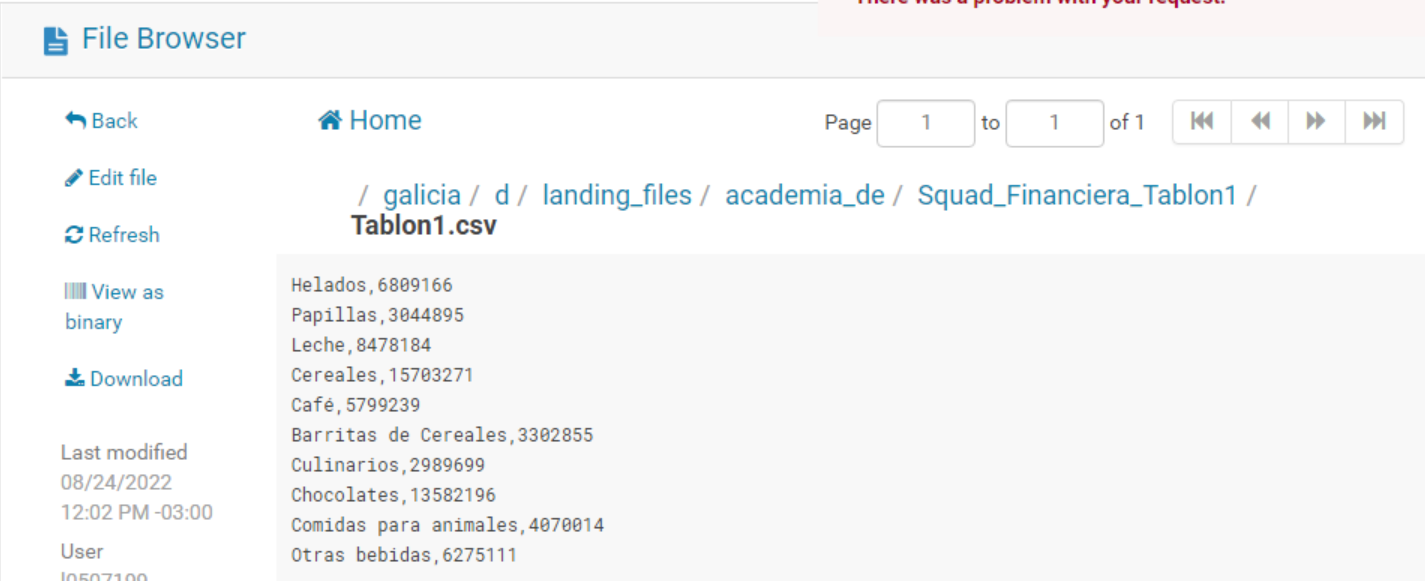


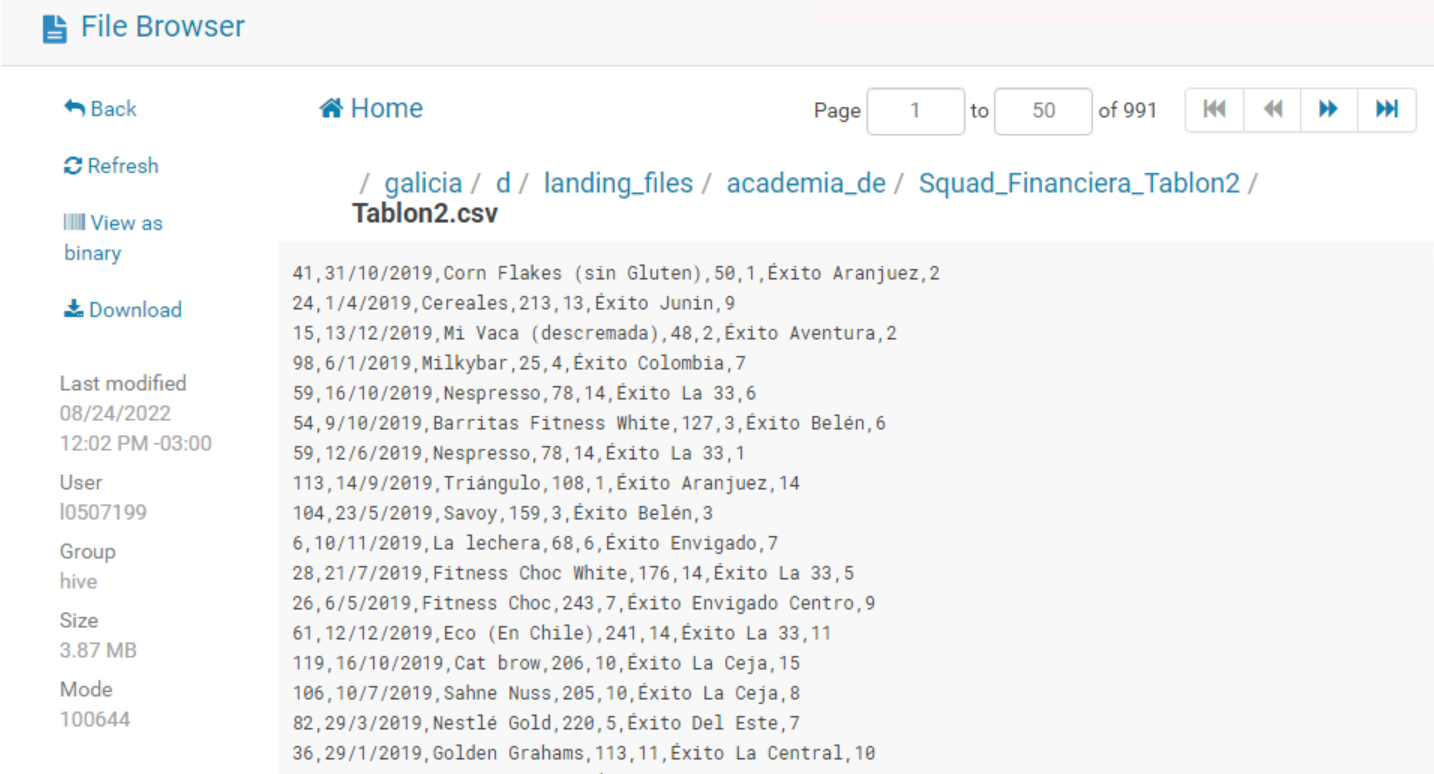






* 21/08/2022 Se completa la documentación de requerimientos y planificación del proyecto.
* 22/08/2022 Se desarrolla la implementación de lógica en tablones y la generación de archivos csv de los mismos.





* 24/08/2022 Se finaliza con la construcción de los 3 reportes.





