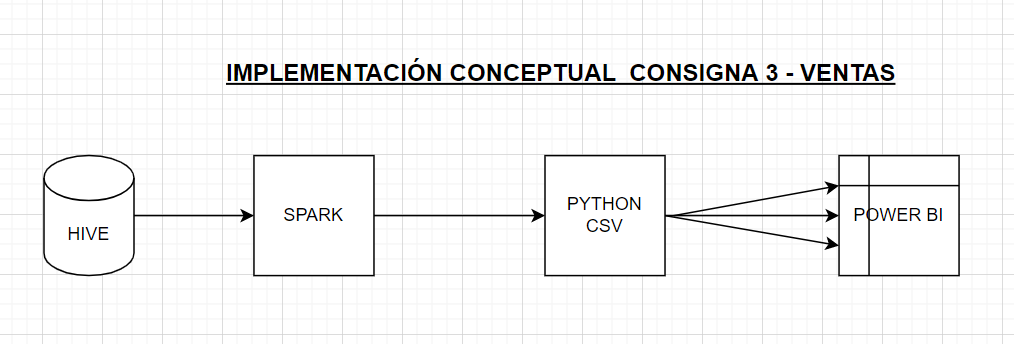
**DOCUMENTACIÓN CONSIGNA NRO 3 – VENTAS**

**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 origen de los datos serán archivos CSV y se utilizaran 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 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 las 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

**TABLONES EN SPARK:**

facturación\_total

top productos

top sucursales

**ARCHIVOS.CSV:**

empleados.csv

fact.csv

sucursales.csv

producto.csv

**POWER BI:**

Reporte.facturación\_total

Reporte.top productos

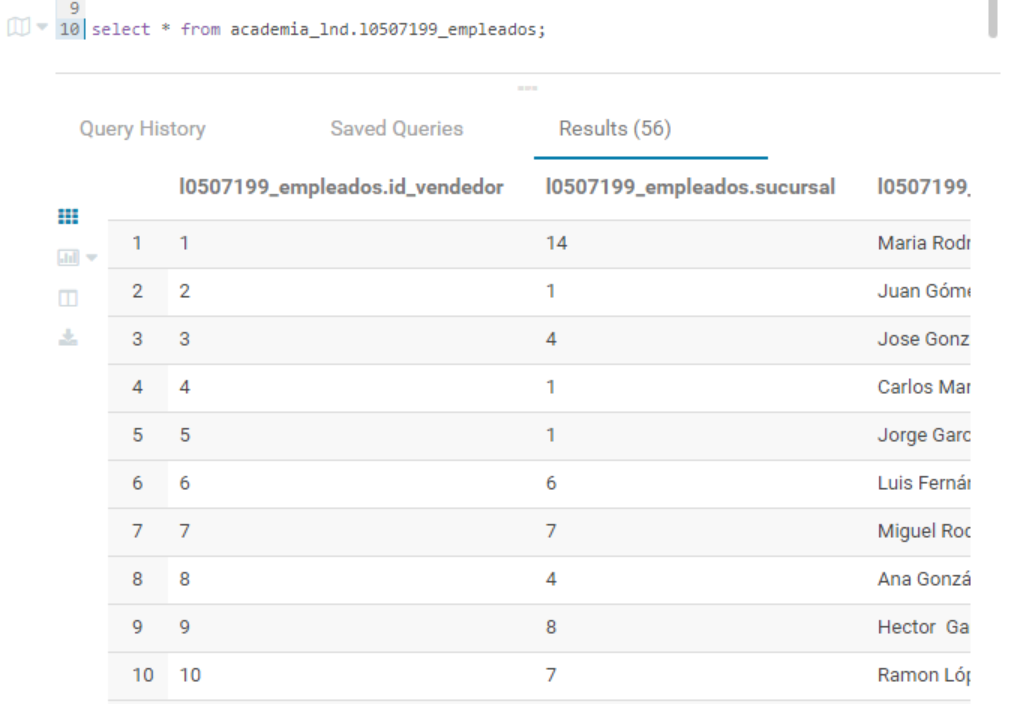
Reporte.top sucursales

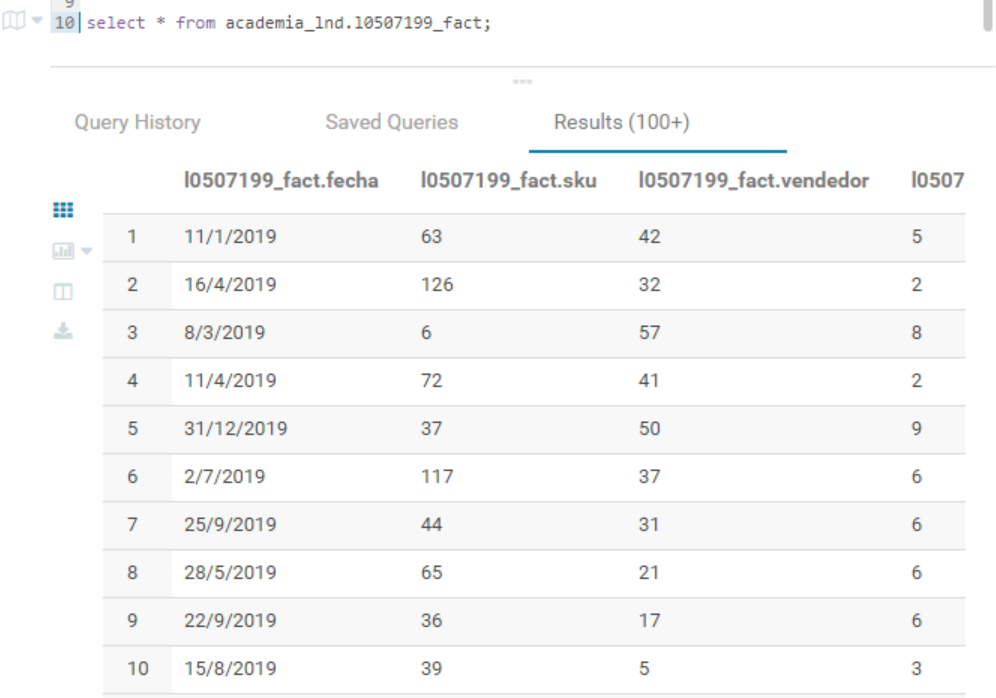
* + - Procesamiento

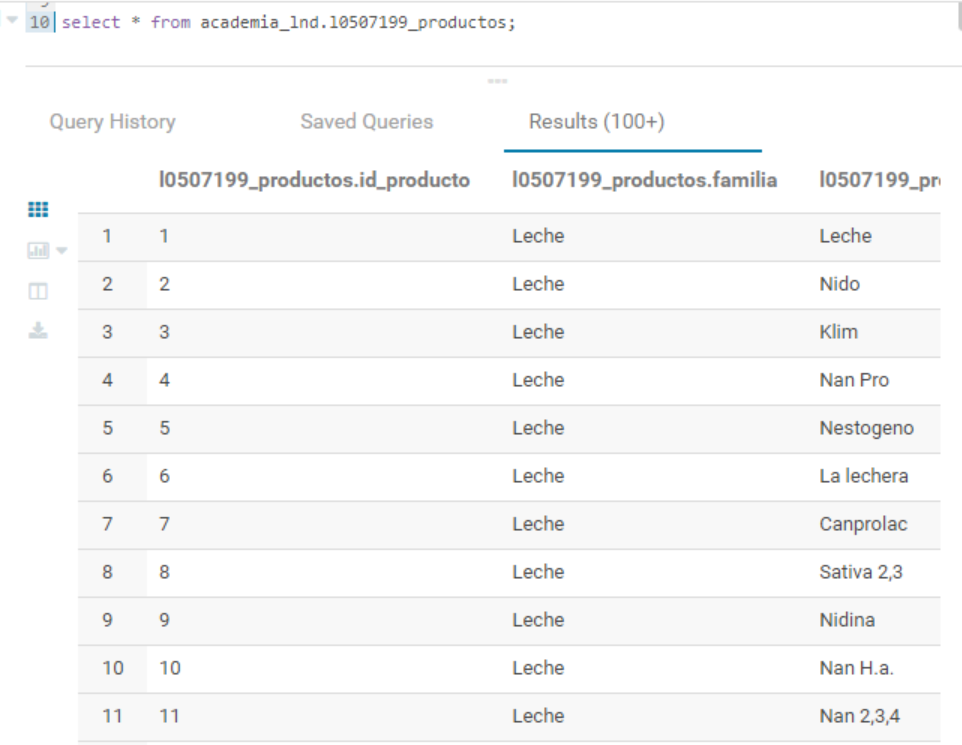
Se consumirán los archivos CSV los cuales serán cargados mediante HIVE en tablas staging con la finalidad de poder realizar la implementación de lógica con Spark, así generando 3 tablones que responderán nuestra consigna relacionada con ventas. Con Python se traspasará la data de los tablones a los archivos CSV para ser tomados finalmente con Power BI y así generar la reportaría correspondiente a nuestro proyecto.

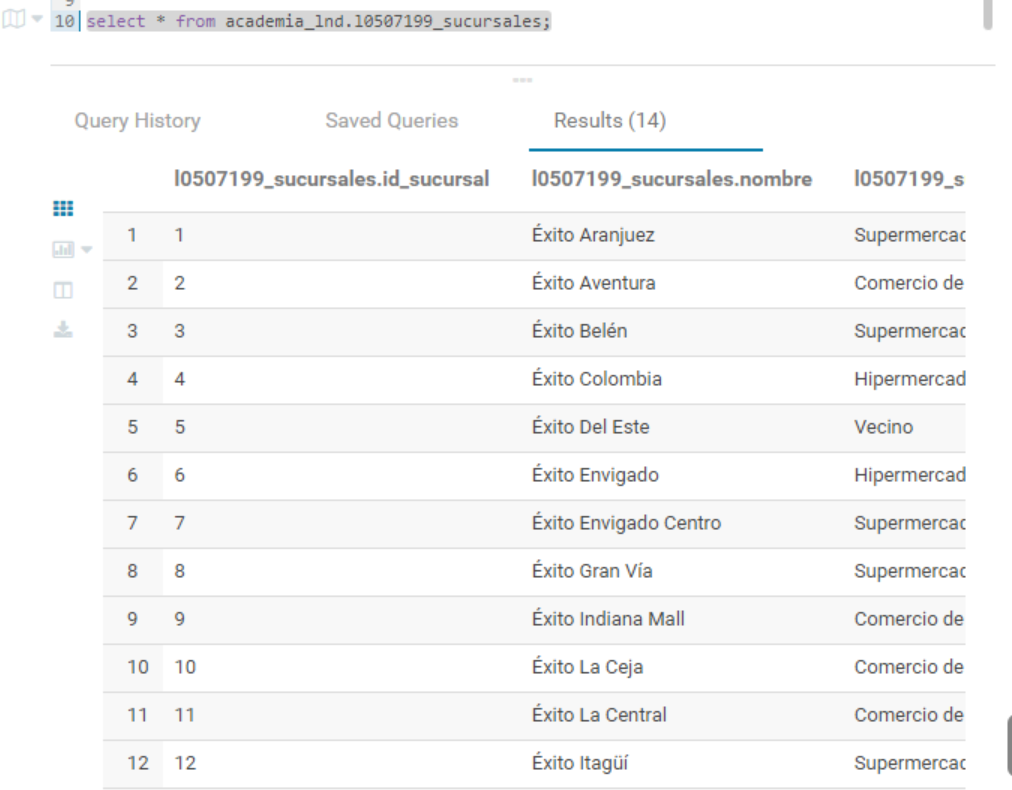
**2. Implementación**

* 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.
* 29/08/2022 Se finaliza con la construcción de los 3 reportes.