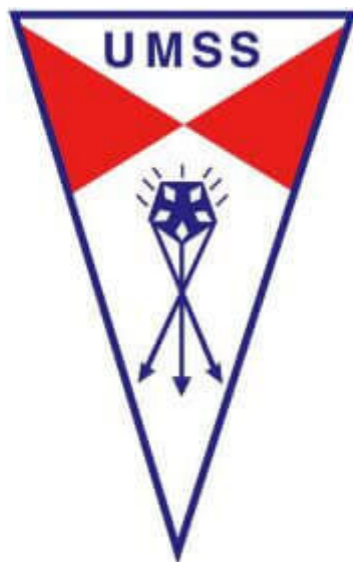


UNIVERSIDAD MAYOR DE SAN SIMÓN

Facultad De Ciencias y Tecnología



PROYECTO PENTAHO

ESTUDIANTES:

- Barra Salaues Dayana Camila
- Lazarte Ordoñez Vanessa Kamar
- Pilco Peralta Mirco Estefano
- Ledezma de la Quintana Kevin Fabricio

DOCENTE: Boris Marcelo Calancha Navia

MATERIA: Business Intelligence - Big Data

30/11/2020

INTRODUCCIÓN.

Los siguientes resultados se obtuvieron de un script de base de datos Excel, teniendo las tablas productos, ventas, vendedor y grupos.

La herramienta PENTAHO nos ayudó a obtener los procesos pedidos, se generó un modelo estrella. Se cargaron datos de entrada a SPOON y se realizaron procesos ETL en MySQL. Con esto se logró el procesamiento de datos para que estos sean más entendibles y así lograr una toma de decisiones más eficiente.

Hoy en día se necesita procesar grandes cantidades de datos de manera más eficiente y rápida, por eso usamos PENTAHO, la cual nos proporciona un análisis rápido de datos al permitirnos mostrar y sumar grandes cantidades de datos.

Se utilizó MySQL como base de datos para conectarse con SPOON, siendo el localhost: 3306

1. COMPLETAR LA TABLA VENTAS CON EL CAMPO TOTAL VENTA (VENTAS.KILOS*PRODUCTOS.PRECIO).

Método usado para el cálculo:

La fórmula que se utilizó en excel para calcular el total de ventas es la siguiente:

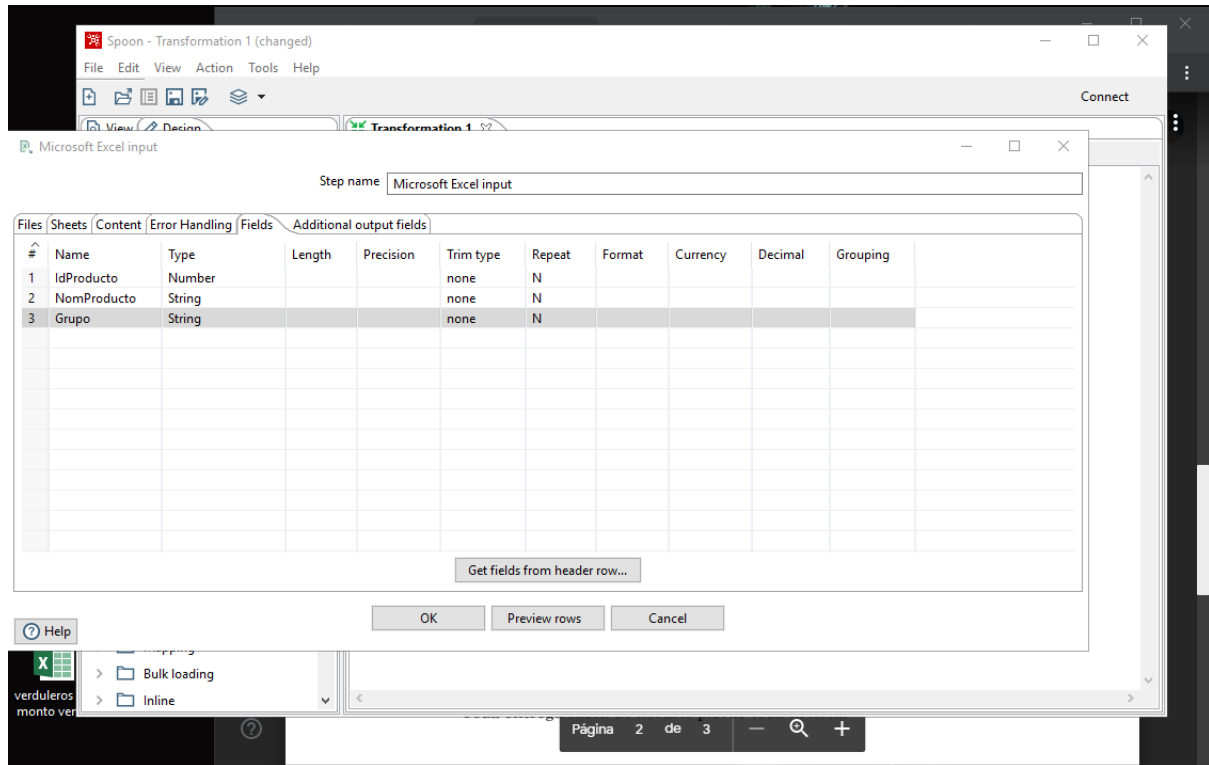
=SI(B3=1,D3*3.9325,0)+SI(B3=2,D3*1.6335,0)+SI(B3=3,D3*1.936,0)+SI(B3=4,D3*0.605,0)+SI(B3=5,D3*2.541,0)+SI(B3=6,D3*2.42,0)+SI(B3=7,D3*0.968,0)+SI(B3=8,D3*3.63,0)+SI(B3=9,D3*1.21,0)+SI(B3=10,D3*0.605,0)+SI(B3=11,D3*1.21,0)+SI(B3=12,D3*2.42,0)+SI(B3=13,D3*0.242,0)+SI(B2=14,D2*3.63,0)

Esta fórmula revisa cual es el código del producto que se encuentra en la columna B y empieza a comparar, dependiendo del valor en la columna B la columna D que es la que contiene los Kilos será multiplicada por el precio del producto. Dicho precio fue sacado de la tabla de producto, la columna precio.

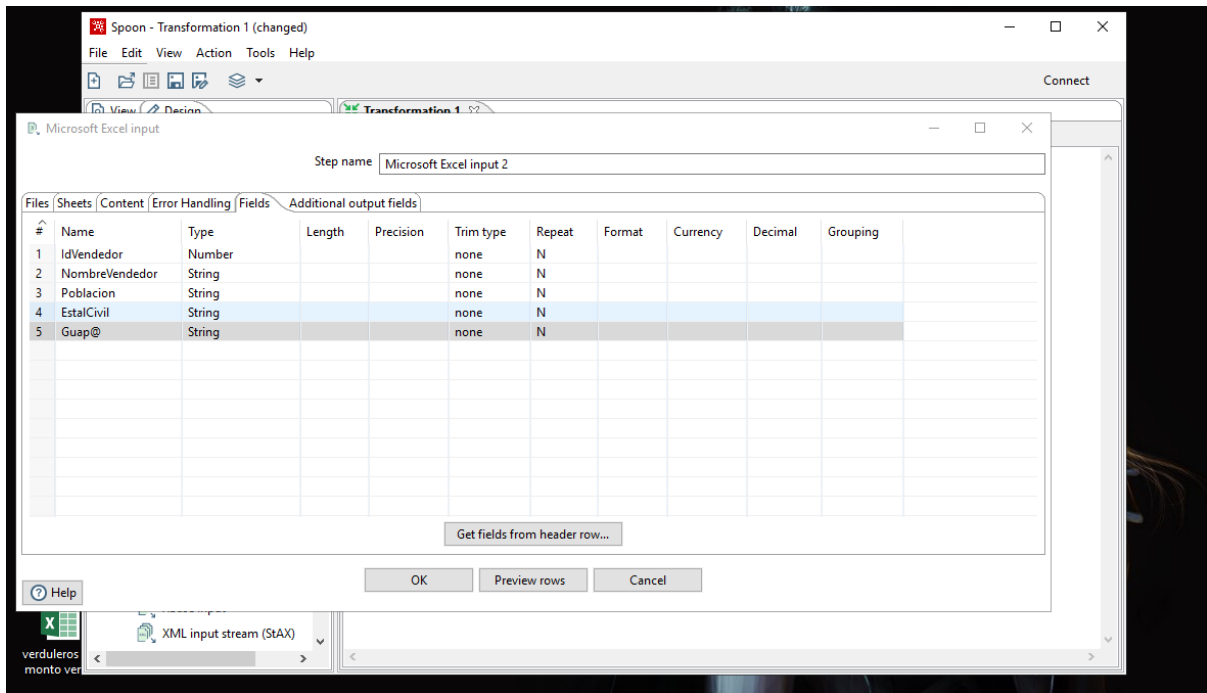
Cod Vendedc	Cod Product	Kilos	TotalVent
4	11	2408	2913,68
4	11	2254	2727,34
4	11	2054	2485,34
4	11	2026	2451,46
4	11	1322	1599,62
4	11	1234	1493,14
4	11	930	1125,3
4	11	636	769,56
4	12	2013	4871,46
4	12	1693	4097,06
4	12	1446	3499,32
4	12	1362	3296,04
4	12	1272	3078,24
4	12	939	2272,38
4	12	502	1214,84
4	9	2148	2599,08
4	9	1511	1828,31
4	9	1387	1678,27
4	7	2076	2009,568
4	7	1896	1835,328

2. REALIZAR EL PROCESO ETL USANDO LA HERRAMIENTA SPOON DE PENTAHO

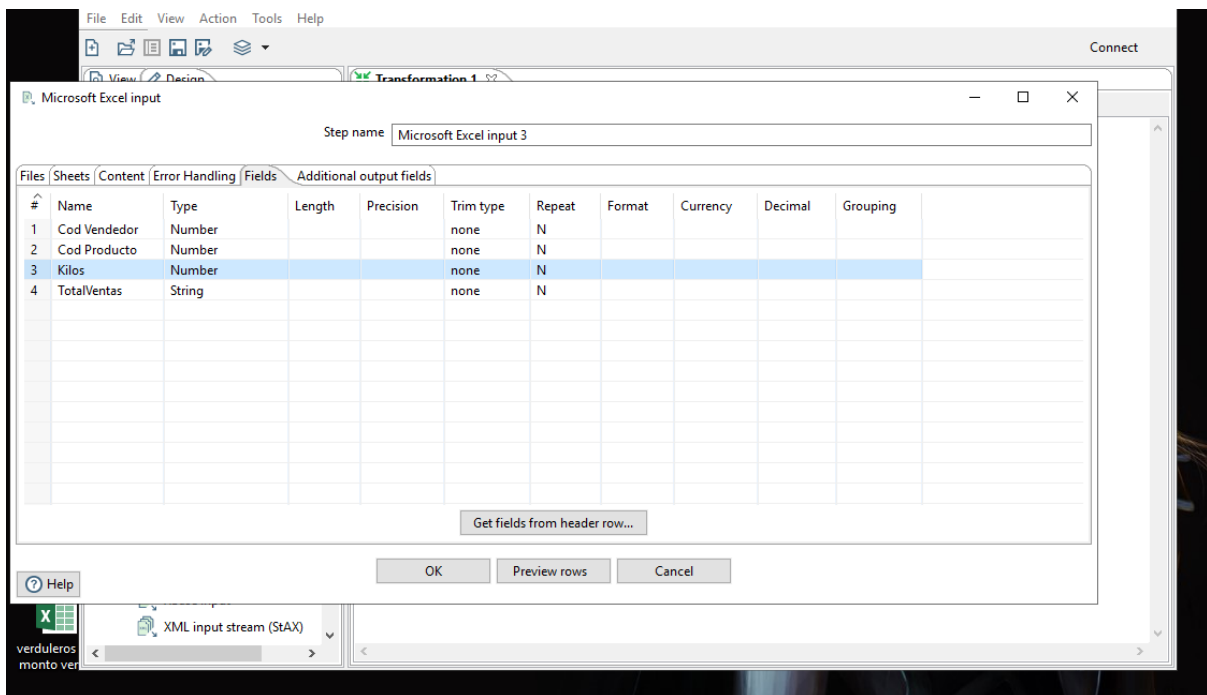
Importación de la hoja Productos



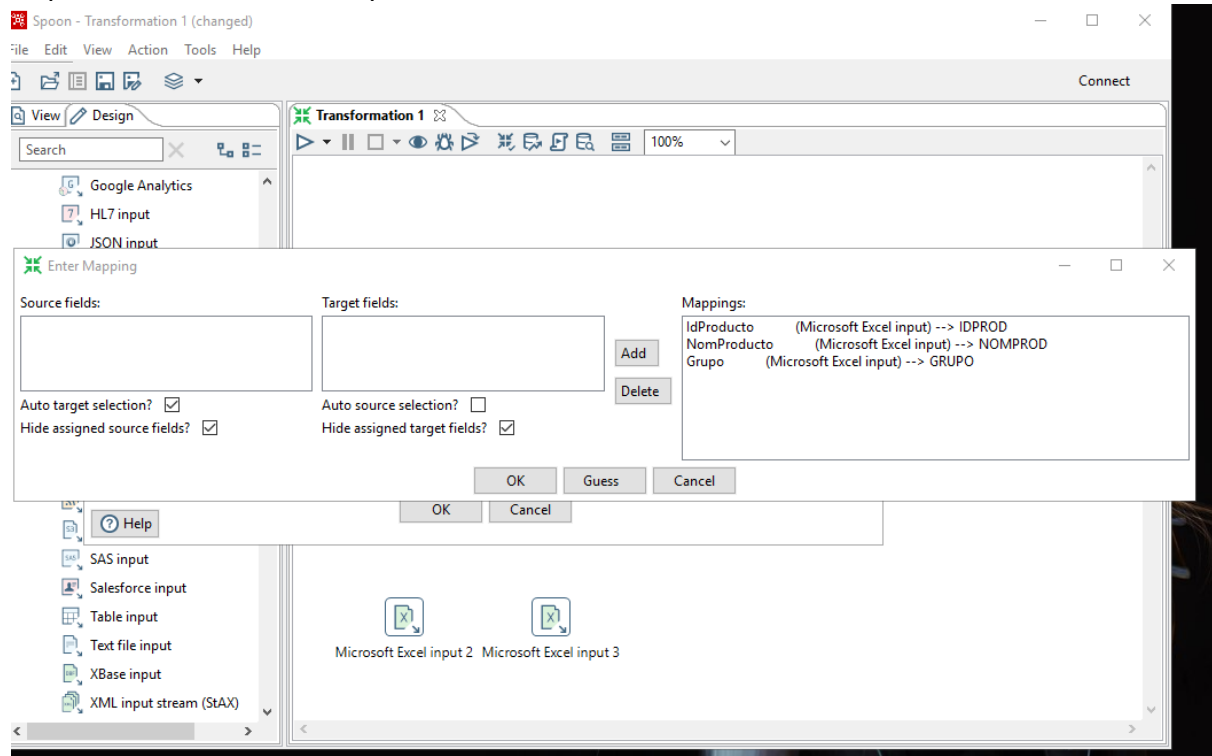
Importación de la hoja vendedores



Importación de la hoja de ventas



mapeo de datos de la tabla producto



The screenshot shows the 'Spoon - Transformation 1 (changed)' window. The 'Table output' step is configured with the following settings:

- Step name: salida
- Connection: verdulerosBD
- Target schema: verduleros
- Target table: productosdim
- Commit size: 1000
- Truncate table: ☐
- Ignore insert errors: ☐
- Specify database fields: ☒

The 'Main options' tab is active, showing the 'Fields to insert' table:

#	Table field	Stream field	
1	IDPROD	IDPROD	
2	NOMPROD	NOMPROD	
3	GRUPO	GRUPO	

The 'Database fields' tab is also visible, showing a table with columns: #, Table field, Stream field, and an empty column. The 'Get fields' button is highlighted in blue.

Me gusta

configuración para la conexión con la base de datos

Database Connection

General
Advanced
Options
Pooling
Clustering

Connection name:
verdulosBD

Connection type:
Ingres VectorWise
Intersystems Cache
KingbaseES
LucidDB
MS Access
MS SQL Server
MS SQL Server (Native)
MariaDB
MaxDB (SAP DB)
MonetDB
MySQL
Native Mondrian
Neoview
Netezza
Oracle
Oracle RDB
Palo MOLAP Server
Pentaho Data Services

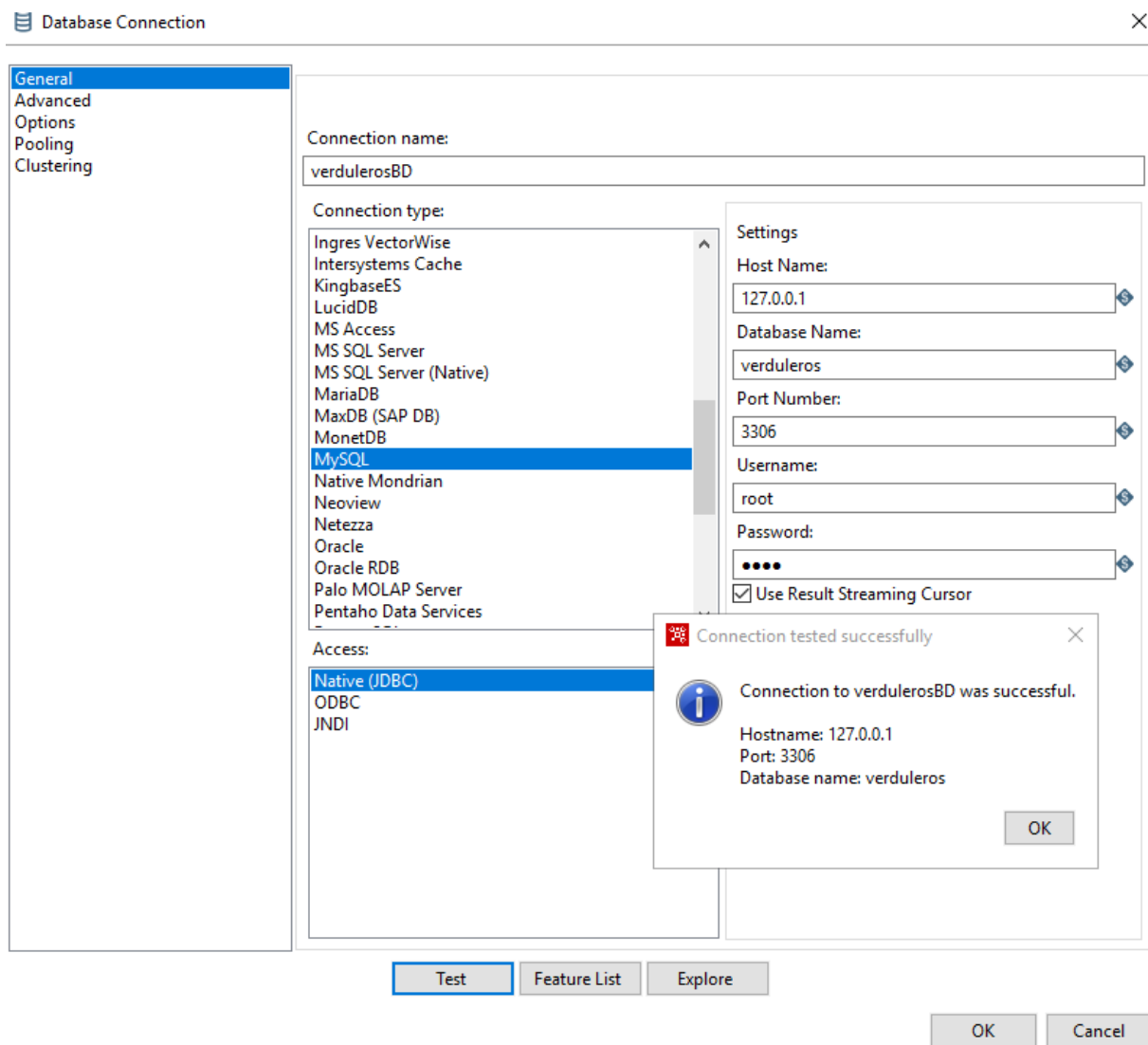
Access:
Native (JDBC)
ODBC
JNDI

Settings
Host Name:
127.0.0.1
Database Name:
verdulos
Port Number:
3306
Username:
root
Password:
.....
☒ Use Result Streaming Cursor

Test Feature List Explore

OK Cancel

prueba de conexión exitosa



proceso ETL exitoso

The screenshot displays the Spoon - corrida1 application window. The interface is divided into several sections:

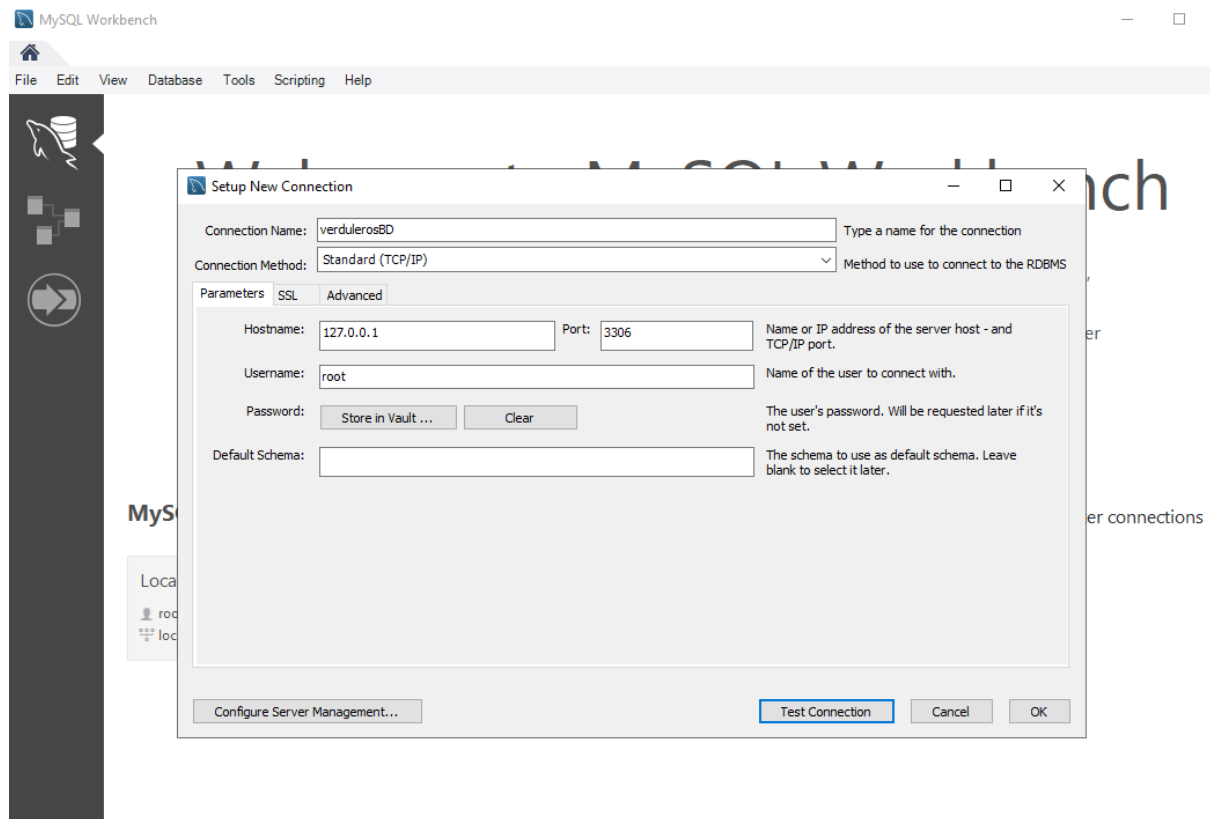
- Top Menu:** File, Edit, View, Action, Tools, Help.
- Left Panel (Design View):** A search bar and a list of input connectors. The "Microsoft Excel input" connector is highlighted.
- Central Canvas:** A workflow diagram showing three steps in sequence: "Microsoft Excel input" (with a green checkmark), "Select values" (with a green checkmark), and "salida" (with a green checkmark). Arrows indicate the flow from left to right.
- Bottom Panel (Execution Results):** A tabbed interface with "Logging" selected. It shows a list of execution logs for the 2021/11/29 15:07:53 session, indicating successful completion for all steps and the overall transformation.

Execution Results Log:

- 2021/11/29 15:07:53 - Spoon - Connected to database [redacted] (commit=1000)
- 2021/11/29 15:07:53 - Microsoft Excel input.0 - Finished processing (I=15, O=0, R=0, W=15, U=0, E=0)
- 2021/11/29 15:07:53 - Select values.0 - Finished processing (I=0, O=0, R=15, W=15, U=0, E=0)
- 2021/11/29 15:07:53 - Microsoft Excel input 3.0 - Finished processing (I=4998, O=0, R=0, W=4998, U=0, E=0)
- 2021/11/29 15:07:53 - salida.0 - Finished processing (I=0, O=15, R=15, W=15, U=0, E=0)
- 2021/11/29 15:07:53 - Microsoft Excel input 2.0 - Finished processing (I=14, O=0, R=0, W=14, U=0, E=0)
- 2021/11/29 15:07:53 - Spoon - The transformation has finished!!

3. MODELAR MODELO ESTRELLA VERDULEROS PARA RESPONDER LA CONSULTA

Creación de la base de datos



script del modelo estrella implementado

```
create database verduleros;
use verduleros;
/*=====*/
/* Table: PRODUCTOSDIM */
/*=====*/
create table PRODUCTOSDIM
(
  IDPROD          integer          not null,
  NOMPROD         varchar(35)      not null,
  GRUPO           varchar(35)      not null,

  constraint PK_PRODUCTOSDIM primary key (IDPROD)
);
/*=====*/
/* Index: PRODUCTOSDIM_PK */
/*=====*/
```

```

/*=====*/
create unique index PRODUCTOSDIM_PK on PRODUCTOSDIM (
IDPROD ASC
);

```

```

/*=====*/
/* Table: VENDEDORESDIM */
/*=====*/
create table VENDEDORESDIM
(
    IDVEND            integer            not null,
    NOMVEND           varchar(35)        null,
    POBLACION         varchar(35)        null,
    ESTADOCIVIL       varchar(35)        null,
    GUAPO             varchar(35)        null,
    constraint PK_VENDEDORESDIM primary key (IDVEND)
);

```

```

/*=====*/
/* Index: VENDEDORESDIM_PK */
/*=====*/
create unique index VENDEDORESDIM_PK on VENDEDORESDIM (
IDVEND ASC
);

```

```

/*=====*/
/* Table: VENTASFACT */
/*=====*/
create table VENTASFACT
(
    IDVEND            integer            null,
    IDPROD            integer            null,
    KILOS             float              null
);

```

```

/*=====*/
/* Index: RELATIONSHIP_1_FK */
/*=====*/
create index RELATIONSHIP_1_FK on VENTASFACT (
IDVEND ASC
);

```

```

/*=====*/
/* Index: RELATIONSHIP_2_FK */
/*=====*/
create index RELATIONSHIP_2_FK on VENTASFACT (
IDPROD ASC
);

```

```

alter table VENTASFACT
    add constraint FK_VENTASFA_RELATIONS_VENDEDOR foreign key (IDVEND)
        references VENDEDORESDIM (IDVEND)
        on update restrict
        on delete restrict;

```

alter table VENTASFACT

add constraint FK_VENTASFA_RELATIONS_PRODUCTO foreign key (IDPROD)
references PRODUCTOSDIM (IDPROD)
on update restrict
on delete restrict;

hn

The screenshot shows a SQL IDE interface. The left sidebar displays a 'SCHEMAS' tree with a search filter 'Filter objects'. Under the 'verduras' schema, the 'Tables' folder is expanded, showing 'productosdim', 'vendedoresdim', and 'ventasfact'. The main query editor shows a script with the following SQL code:

```
16 and ventasfact.IDVEND=14;  
17  
18  
19 • select * from productosdim;  
20 • select * from ventasfact;  
21 • select * from vendedoresdim;  
22
```

Below the query editor, the 'Result Grid' is displayed, showing a table with 3 columns: IDPROD, NOMPROD, and GRUPO. The table contains 8 rows of data:

IDPROD	NOMPROD	GRUPO
1	Mandarinas	Frutas
2	Lechugas	Verduras
3	Melones	Frutas
4	Coles	Verduras
5	Berenjenas	Hortalizas
6	Platanos	Frutas
7	Tomates	Verduras
8	Uvas	Frutas

The bottom of the interface shows an 'Output' section with 'Action Output' selected, and a status bar at the very bottom with columns for '#', 'Time', 'Action', and 'Mensaje'.

toggle automatic h

The screenshot shows a database management interface. On the left, a tree view lists database objects: `vendedoresdim`, `ventasfact`, `Views`, `Stored Procedures`, and `Functions`. Below this, tabs for `Administration` and `Schemas` are visible, with a message `No object selected`. The main area displays a SQL query in a text editor: `select * from vendedoresdim;`. Below the query, a `Result Grid` shows the data. The grid has columns: `IDVEND`, `NOMVEND`, `POBLACION`, `ESTADOCIVIL`, and `GUAPO`. The data is as follows:

IDVEND	NOMVEND	POBLACION	ESTADOCIVIL	GUAPO
1	Pepito	Barcelona	Soltero	VERDADERO
2	Carmen	Madrid	Separado	FALSO
3	Rosa	Martorell	Casado	VERDADERO
4	Gloria	badalona	Divorciado	FALSO
5	Fran	Barcelona	Viudo	VERDADERO
6	Monica	malaga	Arrejado	FALSO
7	Quima	Madrid	Separado	VERDADERO
8	Ramon	Sant Esteve sesrovires	Divorciado	VERDADERO

Below the result grid, there is an `Output` section with a dropdown menu set to `Action Output`. It shows a single action:

#	Time	Action	Message	Duration / Fetch
151	17:12:00	select * from ventasfact LIMIT 0, 1000	1000 row(s) returned	0.000 sec / 0.000 :

At the bottom left, there is a tab for `Object Info` with a sub-tab for `Session`.

Las ventas totalizadas en kilos / montos de dinero, asociados a los vendedores clasificados por la población donde residen, asociado también a su estado civil, que sean o no sean guapos y que vendieron entre los tipos de Productos frutas y/o verduras.

Consulta SQL

```
select distinct(NOMVEND),sum(kilos),sum(monto),ESTADOCIVIL,GUAPO,POBLACION
from vendedoresdim,ventasfact,productosdim
where vendedoresdim.IDVEND=ventasfact.IDVEND
and ventasfact.IDPROD=productosdim.IDPROD
and not GRUPO='Hortalizas'
group by(NOMVEND);
```

MySQL Workbench

data x unconnected x

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

- sys
- verduras
 - Tables
 - productosdim
 - vendedoresdim
 - ventasfact
 - Views
 - Stored Procedures
 - Functions

Administration Schemas

Information

No object selected

script verduleros actualizado consulta x SQL File 2* SQL File 3

Limit to 1000 rows

```
1 • select distinct(NOMVEND),sum(kilos),sum(monto),ESTADOCIVIL,GUAPO
2 from vendedoresdim,ventasfact,productosdim
3 where vendedoresdim.IDVEND=ventasfact.IDVEND
4 and ventasfact.IDPROD=productosdim.IDPROD
5 and not GRUPO='Hortalizas'
6 group by(NOMVEND);
7
```

Result Grid

	NOMVEND	sum(kilos)	sum(monto)	ESTADOCIVIL	GUAPO
▶	Gloria	421295	937428	Divorciado	FALSO
	Carla	402526	881924	Arrejado	VERDADERO
	Pepito	385143	858262	Soltero	VERDADERO
	Fran	413720	946197	Viudo	VERDADERO
	Carmen	412981	914728	Separado	FALSO
	Monica	396125	871160	Arrejado	FALSO
	Rosa	393611	904137	Casado	VERDADERO
	Federico	354405	807643	Casado	VERDADERO

Result 95 x

Read Only Context Help Snippets

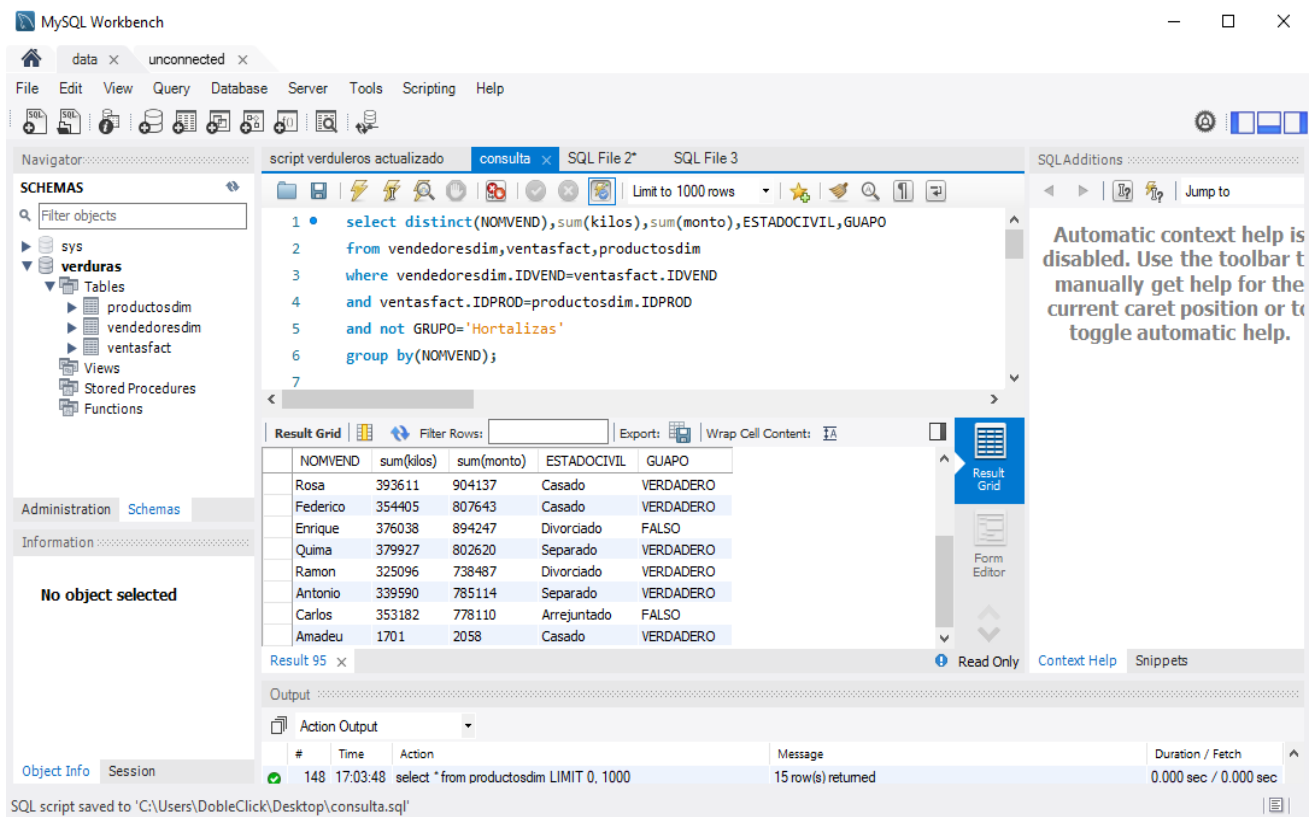
Output

Action Output

#	Time	Action	Message	Duration / Fetch
148	17:03:48	select * from productosdim LIMIT 0, 1000	15 row(s) returned	0.000 sec / 0.000 sec

SQL script saved to 'C:\Users\DobleClick\Desktop\consulta.sql'

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.



CONCLUSIONES.

Para la realización de este proyecto se usó las herramientas de Pentaho, Se cargó la base de datos VERDULEROS a la herramienta spoon para obtener los datos de entrada y así realizar el proceso ETL en el modelo estrella en MySQL.

Donde obtuvimos los siguientes resultados:

El monto total de la venta de frutas fue de 9.573.695 bolivianos, mientras que el monto total de la venta de verduras fue de 1548420 bolivianos. Representando que la venta de frutas es mucho mayor que la de verduras, además siendo esta la que mayor ganancia genera.

La población con mayor adquisición de productos es: Sant Esteve Sesrovires con 1.40 millones de kilos y 3.23 millones de bolivianos, seguido de Madrid con 1.15 millones de kilos y 2.5 millones de bolivianos.

La población con menor cantidad de ingresos y ventas es: Barcelona con 858.62 mil bolivianos y 385.14 mil kilos vendidos.

Entre los motivos de porque hay pocos ingresos en esta población, estaría la poca cantidad de vendedores en la población (4 vendedores en Sant Esteve Sesrovires y tan solo uno para Barcelona).

La vendedora que mayores ventas de frutas y verduras ha realizado por su población es Gloria en Badalona con 421.295 kilos y 937.44 mil bolivianos, de características: divorciada

y no guapa.

En cambio, el que menores ventas realizó en su población es Amadeu en Sant Esteve Sesrovires con 1701 kilos y 2058 bolivianos con las características: guapo y casado.

Concluimos mencionando que Pentaho ofrece un rendimiento eficaz dentro de un mercado competitivo. Ya que ofrece en su funcionalidad la implementación en una organización, el cual nos permite interactuar, a través de los datos, la toma de decisiones, estableciendo una evaluación en la ponderación de un resultado.

La optimización en la funcionalidad de una herramienta como Pentaho facilita los procesos como un factor estratégico para identificar oportunidades para generar estrategias e implementar la información obtenida en la toma de decisiones, para dar soluciones en el mercado de negocios, para definir el control de gastos de recursos y el monitoreo, análisis y administración de la información. Pentaho es una solución muy potente para el desarrollo de proyectos BI, aunque algunas veces presenta inestabilidades con los plugins CE, otro aspecto muy importante a tomar en cuenta es que Pentaho es código abierto.

	NOMVEND	sum(kilos)	sum(monto)	ESTADOCIVIL	GUAPO	POBLACION
►	Fran	413720	946197	Viudo	VERDADERO	Barcelona
	Gloria	421295	937428	Divorciado	FALSO	badalona
	Carmen	412981	914728	Separado	FALSO	Madrid
	Rosa	393611	904137	Casado	VERDADERO	Martorell
	Enrique	376038	894247	Divorciado	FALSO	Sant Esteve sesrovires
	Carla	402526	881924	Arreuntado	VERDADERO	La Beguda Alta
	Monica	396125	871160	Arreuntado	FALSO	malaga
	Pepito	385143	858262	Soltero	VERDADERO	Barcelona
	Federico	354405	807643	Casado	VERDADERO	Sant Esteve sesrovires
	Quima	379927	802620	Separado	VERDADERO	Madrid
	Antonio	339590	785114	Separado	VERDADERO	Sant Esteve sesrovires
	Carlos	353182	778110	Arreuntado	FALSO	Madrid
	Ramon	325096	738487	Divorciado	VERDADERO	Sant Esteve sesrovires
	Amadeu	1701	2058	Casado	VERDADERO	Sant Esteve sesrovires

monto total de venta de Frutas

Result Grid		Filter Rows:		Export:		Wrap Cell Content:	
	sum(monto)						
▶	9573695						

Result 109 x

monto total de venta de Verduras

	sum(monto)
▶	1548420

mayor venta de Frutas que de verduras