

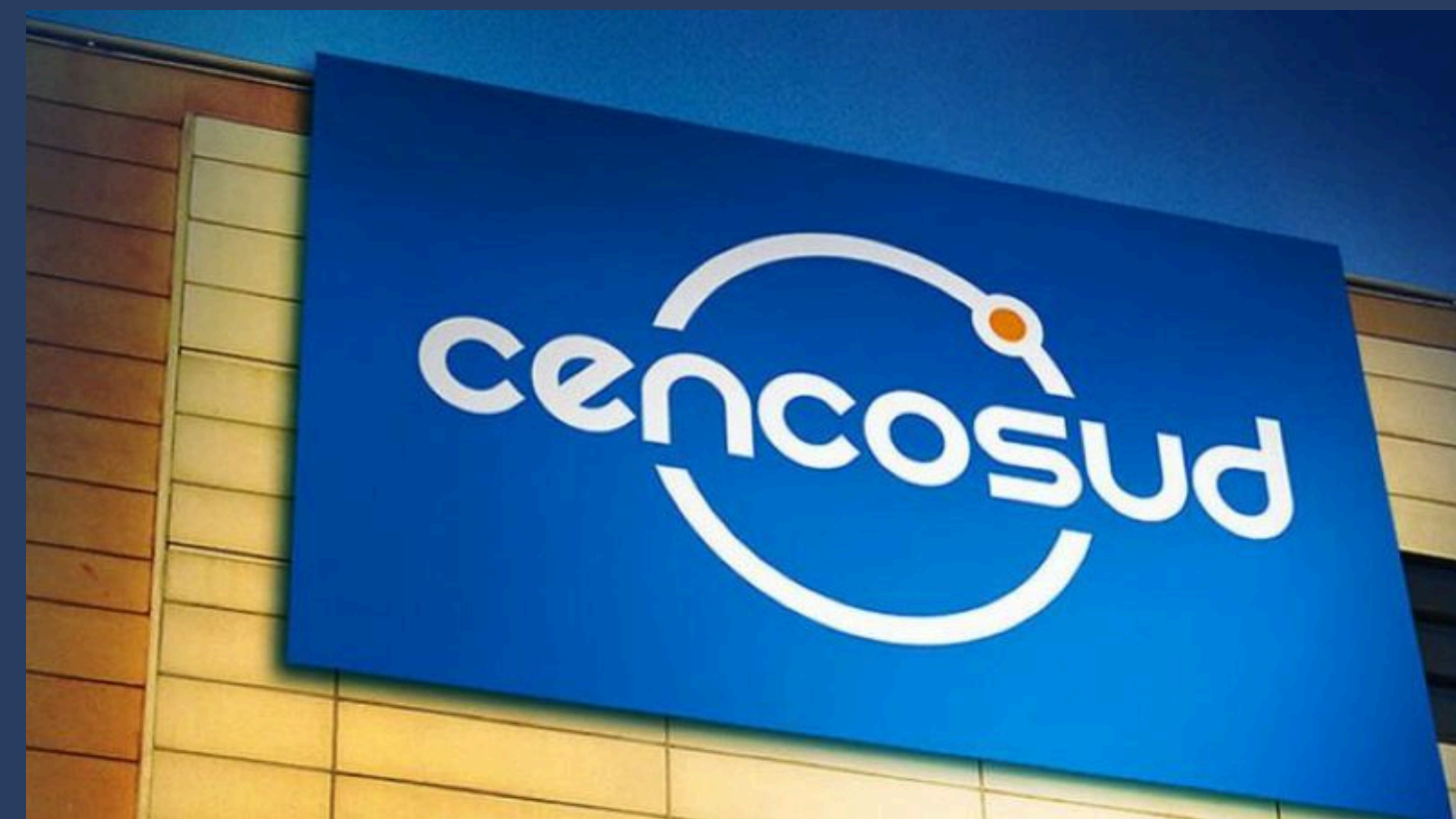
CENCOSUD



Curso: Sistema de Inteligencia de Negocio

Profesor: Dr. Ing. Aradiel Castañeda, Hilario

Alumnos:
Larico Cruz, Diego César - 20212679A
Cabana Cazani, Gabriel Alessandro -



PREGUNTAS DEL NEGOCIO



Área	Rol	Pregunta de negocio	Prioridad	Fuente de datos
CRM	Analista de Clientes	¿Qué clientes presentan mayor o menor recencia de compra y cómo se segmentan por tipo de cliente?	Alta	Hecho_Venta + Cliente + Periodo
CRM	Gerente de Clientes	¿Cuál es la frecuencia promedio de compra por cliente y categoría de producto?	Alta	Hecho_Venta + Cliente + Producto
CRM	Gerente de Clientes	¿Cuál es el monto promedio de compra por cliente en las distintas cadenas (Wong, Metro, etc.)?	Alta	Hecho_Venta + Cliente + TiendaCanal
Gobierno de Datos	Data Steward	¿Cuál es el ticket promedio por transacción y cómo varía por canal y tienda?	Alta	Hecho_Venta + TiendaCanal
Gobierno de Datos	Analista Financiero	¿Cuál es el margen promedio por ticket y qué productos aportan más a la rentabilidad?	Alta	Hecho_Venta + Producto
Ecommerce	Jefe de Promociones	¿Qué porcentaje de la venta bruta corresponde a descuentos aplicados en promociones?	Media	Hecho_Venta + Promocion_precio
Ecommerce	Jefe de Promociones	¿Qué porcentaje de tickets tienen promociones y cuál es su impacto en el volumen de ventas?	Media	Hecho_Venta + Promocion_precio
Ecommerce	Category Manager	¿Cuál es la venta neta por categoría y subcategoría de producto?	Alta	Hecho_Venta + Producto
Calidad de Datos	Auditor de Datos	¿Cuál es la participación por canal de venta (Wong vs. Metro u otros) y cómo se distribuye por tienda?	Alta	Hecho_Venta + TiendaCanal
Calidad de Datos	Auditor de Datos	¿Qué productos o promociones no tienen información de vigencia o están mal registrados?	Alta	Producto + Promocion_precio

IDENTIFICACION DE KPI'S

Nombre KPI	Descripción	Fórmula	Unidad	Periodicidad	Fuente de Datos	Responsable
Ticket Promedio	Mide el gasto promedio por transacción, desagregado por canal y tienda	$\Sigma(\text{monto_venta_neta}) \div \text{N}^\circ \text{ tickets}$	Soles	Diario	Hecho_Venta + TiendaCanal	Gobierno de Datos
Margen Promedio por Ticket	Evalúa la rentabilidad de cada ticket considerando productos vendidos	$\Sigma(\text{monto_margen}) \div \text{N}^\circ \text{ tickets}$	Soles	Diario	Hecho_Venta + Producto	Gobierno de Datos
% Descuento sobre Venta Bruta	Mide el impacto de la política de promociones en la venta bruta	$(\Sigma(\text{monto_descuento}) \div \Sigma(\text{monto_venta_bruta})) \times 100$	%	Diario/Mensual	Hecho_Venta + Promocion_precio	Ecommerce
% de Tickets con Promoción	Mide el alcance de las promociones en el total de tickets	$(\text{N}^\circ \text{ tickets con cod_promoción} \div \text{N}^\circ \text{ total tickets}) \times 100$	%	Mensual	Hecho_Venta + Promocion_precio	Ecommerce
Venta Neta por Categoría/Subcategoría	Identifica qué productos generan más ingresos en cada categoría	$\Sigma(\text{monto_venta_neta}) \text{ agrupado por categoría/subcategoría}$	Soles	Mensual	Hecho_Venta + Producto	Ecommerce
Frecuencia Promedio de Compra	Mide cuántas compras realiza un cliente promedio en un periodo	$\Sigma(\text{trx_acum}) \div \text{N}^\circ \text{ clientes}$	Compras/Cliente	Mensual	Hecho_Venta + Cliente + Periodo	CRM
Monto Promedio de Compra por Cliente	Evalúa cuánto gasta un cliente promedio considerando canales y tiendas	$\Sigma(\text{venta_acum}) \div \text{N}^\circ \text{ clientes}$	Soles	Mensual	Hecho_Venta + Cliente + TiendaCanal	CRM
Recencia Promedio de Compra	Identifica cada cuántos días vuelven a comprar los clientes	$\text{AVG}(\text{recencia_días})$	Días	Mensual	Hecho_Venta + Cliente + Periodo	CRM
Visitas Promedio por Cliente	Mide la intensidad de visitas de los clientes	$\Sigma(\text{visitas_mes}) \div \text{N}^\circ \text{ clientes}$	Visitas/Cliente	Mensual	Hecho_Venta + Cliente	CRM
Participación de Canal	Compara el peso de cada canal (Wong, Metro, etc.) en las ventas totales	$(\Sigma(\text{ventas_canal}) \div \Sigma(\text{ventas_totales})) \times 100$	%	Mensual	Hecho_Venta + TiendaCanal	Calidad de Datos
Calidad de Registro de Productos/Promociones	Evalúa consistencia de datos en productos y promociones (vigencias, categorías, codificación)	% de registros con atributos completos	%	Mensual	Producto + Promocion_precio	Calidad de Datos

MODELO CONCEPTUAL

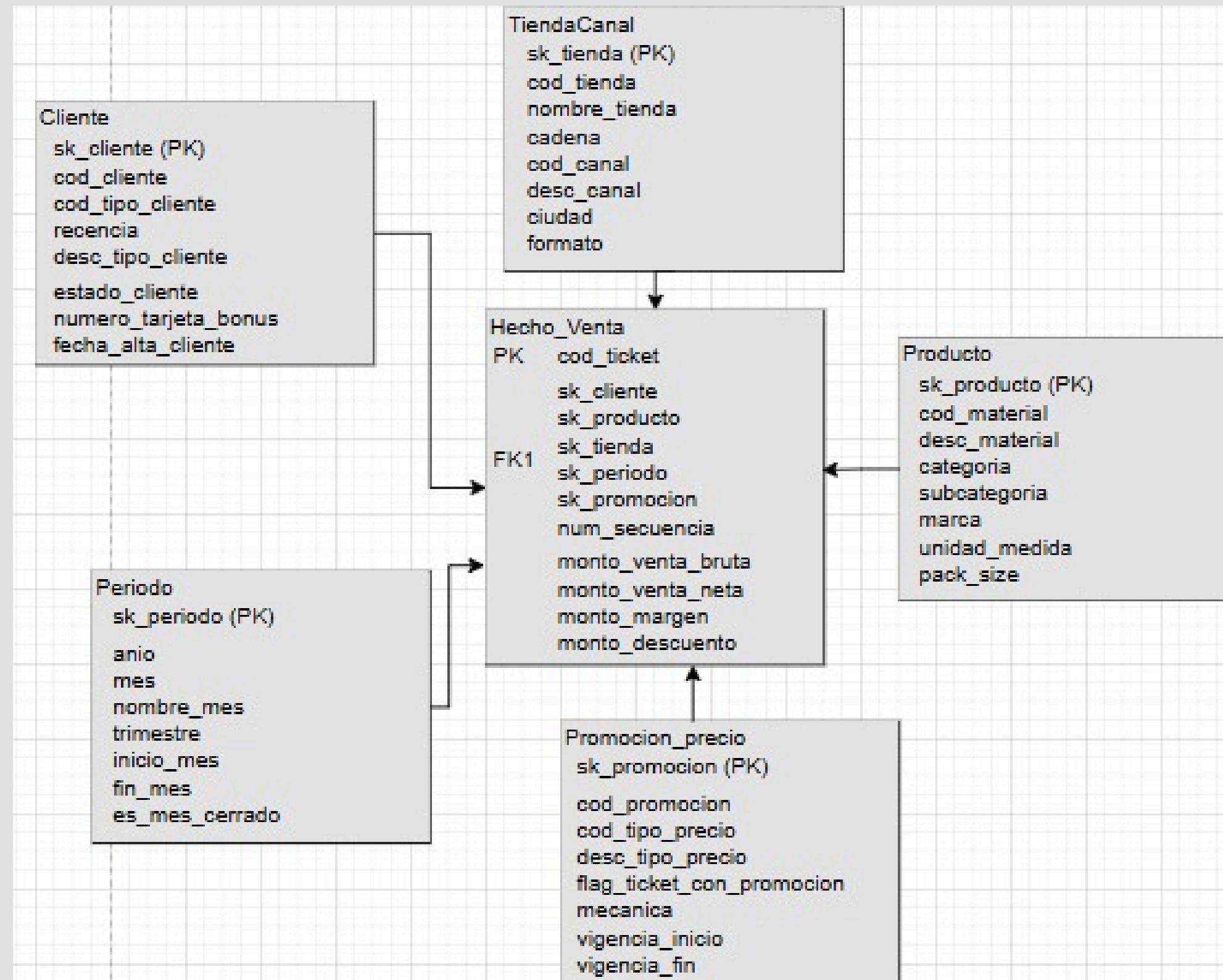


TABLA DE HECHOS

Hecho	Métricas principales
<u>Hecho Venta</u>	<ul style="list-style-type: none">- <u>monto_venta_bruta</u> (valor total antes de descuentos)- <u>monto_venta_neta</u> (valor después de descuentos)- <u>monto_margen</u> (rentabilidad por <u>ticket</u>)- <u>monto_descuento</u> (descuentos aplicados)- <u>num_secuencia</u> (orden del <u>ticket</u>/venta)

TABLA DE DIMENSIONES

Dimensión	Jerarquía / Atributos	Justificación
Periodo	Año → Trimestre → Mes → Día	Permite analizar las ventas y KPIs con una perspectiva temporal(comparación mensual, trimestral y anual).
Cliente	Tipo cliente → Segmento → Subsegmento → Cliente individual	Facilita el análisis de comportamiento de compra, recencia, frecuencia y valor del cliente (RFM).
Producto	Categoría → Subcategoría → Marca → SKU / Pack Size	Permite evaluar la venta neta y márgenes por categoría, subcategoría y producto específico.
TiendaCanal	Cadena (Wong, Metro) → Canal → Ciudad → Tienda → Formato	Brinda análisis de participación por canal, región y punto de venta.
Promoción	Tipo promoción → Mecánica → Vigencia (inicio-fin) → Promoción específica	Permite medir el impacto de descuentos y promociones en ventas y márgenes.

ANALISIS DE SISTEMAS OLTP

Sistema	Área usuaria	Tipo	Tecnología	Frecuencia actualización	Observaciones
Sistema de Ventas (POS)	Gobierno de Datos / Finanzas	Transaccional de ventas por ticket	Redshift / SQL Server (POS)	Diario / Tiempo real	Fuente principal para la tabla <u>Hecho_Venta</u> (monto_venta_bruta, monto_venta_neta, márgenes, descuentos).
Sistema Maestro de Clientes (CRM)	CRM	Maestro de clientes	Redshift / SAP CRM	Semanal	Contiene <u>cod_cliente</u> segmentado, RFM, número de tarjeta; se integra con las ventas para análisis de recencia y frecuencia.
Sistema de Tiendas y Canales	Calidad de Datos / Comercial	Maestro organizacional	Redshift Retail / SAP	Diario	Define sucursales, cadenas (Wong, Metro), canal y formato. Permite calcular participación de canal.
Sistema de Productos (Catálogo)	E-commerce / Comercial	Maestro de productos	SAP Retail / SQL Server	Diario / Semanal	Registra categorías, subcategorías, marcas, unidad de medida. Base para el KPI de venta neta por producto.
Sistema de Promociones	E-commerce	Transaccional de promociones y descuentos	Redshift / SAP	Diario	Contiene promociones activas, vigencia, mecánicas y tipo de descuento; necesario para KPIs de % descuentos y % tickets con promoción.

PARTE PRACTICA



INGESTA DE DATOS Nos localizamos en Files view para cargar los archivos csv y almacenarlos en la carpeta Data/raw

Ambari

Sandbox

1 op

3 alerts

DashboardServicesHostsAlertsAdmin

raj_ops

Total: 12 files or folders

+ Select All

New Folder

Upload

3

Search in current directory...

Name >	Size >	Last Modified >	Owner >	Group >	Permission
app-logs	--	2018-06-18 10:18	yarn	hadoop	drwxrwxrwx
apps	--	2018-06-18 11:13	hdfs	hdfs	drwxr-xr-x
ats	--	2018-06-18 09:52	yarn	hadoop	drwxr-xr-x
data	--	2025-09-29 16:03	raj_ops	hdfs	drwxr-xr-x
hdp	--	2018-06-18 09:52	hdfs	hdfs	drwxr-xr-x
livy2-recovery	--	2018-06-18 10:11	livy	hdfs	drwx-----
mapred	--	2018-06-18 09:52	mapred	hdfs	drwxr-xr-x
mr-history	--	2018-06-18 09:52	mapred	hadoop	drwxrwxrwx
ranger	--	2018-06-18 10:59	hdfs	hdfs	drwxr-xr-x

Ambari

Sandbox

0 ops

3 alerts

DashboardServicesHostsAlertsAdmin

raj_ops

Total: 1 files or folders

+ Select All

New Folder

Upload

3

Search in current directory...

Name >	Size >	Last Modified >	Owner >	Group >	Permission
raw	--	2025-09-29 16:05	raj_ops	hdfs	drwxr-xr-x

Usamos el siguiente comando para cargar el archivo csv al HDFS

```
[root@sandbox-hdp ~]# hdfs dfs -ls /
Found 12 items
drwxrwxrwx - yarn hadoop 0 2018-06-18 15:18 /app-logs
drwxr-xr-x - hdfs hdfs 0 2018-06-18 16:13 /apps
drwxr-xr-x - yarn hadoop 0 2018-06-18 14:52 /ats
drwxr-xr-x - raj_ops hdfs 0 2025-09-29 21:03 /data
drwxr-xr-x - hdfs hdfs 0 2018-06-18 14:52 /hdp
drwx----- - livy hdfs 0 2018-06-18 15:11 /livy2-recovery
drwxr-xr-x - mapred hdfs 0 2018-06-18 14:52 /mapred
drwxrwxrwx - mapred hadoop 0 2018-06-18 14:52 /mr-history
drwxr-xr-x - hdfs hdfs 0 2018-06-18 15:59 /ranger
drwxrwxrwx - spark hadoop 0 2025-09-29 22:36 /spark2-history
drwxrwxrwx - hdfs hdfs 0 2018-06-18 16:06 /tmp
drwxr-xr-x - hdfs hdfs 0 2018-06-18 16:08 /user
[root@sandbox-hdp ~]# hdfs dfs -ls /data
Found 1 items
drwxr-xr-x - raj_ops hdfs 0 2025-09-29 21:05 /data/raw
[root@sandbox-hdp ~]# hdfs dfs -put dim_cliente /data/raw
```


Vemos los archivos cargados en el HDFS

Ambari Sandbox0 ops2 alerts

DashboardServicesHostsAlertsAdmin

raj_ops

HomeRefreshRecycle

/ > data > raw

Total: 7 files or folders

Select AllNew FolderUpload

3

Search in current directory...

Q

Name >	Size >	Last Modified >	Owner >	Group >	Permission
↶					
dim_cliente.csv	5.9 kB	2025-09-29 16:04	raj_ops	hdfs	-rw-r--r--
dim_cliente_rfm.csv	4.1 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_periodo_mensual.csv	4.4 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_producto.csv	6.6 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_promocion_precio.csv	5.3 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
dim_tienda_canal.csv	5.6 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--
tabla_salida.csv	5.2 kB	2025-09-29 16:05	raj_ops	hdfs	-rw-r--r--

Cargamos la tabla para las consultas sql

Ambari Sandbox0 ops0 alerts

DashboardServicesHostsAlertsAdmin

raj_ops

HiveQuerySaved QueriesHistoryUDFsUpload Table

Upload Progress

- Successfully created Actual table.
- Waiting for creation of Temporary table.

Upload from Local

File typeCSV

Databasedefault

Stored asORC

Table nameTrx_ticket

Contains endlines?☐

Upload Table

cod_ticketnum_secuenciask_clientesk_producto

STRINGINTINTINT

TK0001	2	52	57
TK0002	3	47	41
TK0003	2	66	54
TK0004	1	53	24

Realizamos consulta básica para ver los 10 primeros registros

Query Editor

Worksheet * x

trx_ticket sample x

1 SELECT * FROM trx_ticket LIMIT 10;

Query Process Results (Status: SUCCEEDED)

Save results...

LogsResults

Filter columns...

previousnext

trx_ticket.cod_ticket	trx_ticket.num_secuencia	trx_ticket.sk_cliente	trx_ticket.sk_producto	trx_ticket.sk_tienda	trx_ticl
TK0001	2	52	57	99	76
TK0002	3	47	41	47	40
TK0003	2	66	54	81	68
TK0004	1	53	24	8	73
TK0005	4	89	53	2	40
TK0006	3	13	76	2	4
TK0007	5	83	69	66	74
TK0008	4	98	66	14	13
TK0009	2	79	56	8	2
TK0010	1	46	36	22	35

Usamos lo siguiente para poder crear la tabla externa

Query Editor

Worksheet

```
1 CREATE EXTERNAL TABLE IF NOT EXISTS tabla_salida_external (  
2   cod_ticket STRING,  
3   cod_cliente INT,  
4   cod_sucursal INT,  
5   cod_fecha INT,  
6   mto_venta_bruta DECIMAL(18,2),  
7   mto_descuento DECIMAL(18,2),  
8   mto_venta_neta DECIMAL(18,2),  
9   mto_margen DECIMAL(18,2)  
10 )  
11 ROW FORMAT DELIMITED  
12 FIELDS TERMINATED BY '\t'  
13 STORED AS TEXTFILE  
14 LOCATION '/data/raw';
```

SQL

⚙️

📈

🔗

TEZ

✉️

Actualizamos y si se creó la tabla

Database Explorer

default

Search tables...

Databases

default

▢tabla salida external

▢trx ticket

foodmart

Vemos el formato de nuestra tabla

Query Editor

Worksheet

```
1 describe formatted tabla_salida_external;
```

col_name	data_type	comment
# col_name	data_type	comment
""	null	null
cod_ticket	string	""
cod_cliente	int	""
cod_sucursal	int	""
cod_fecha	int	""
mto_venta_bruta	decimal(18,2)	""
mto_descuento	decimal(18,2)	""
mto_venta_neta	decimal(18,2)	""
mto_margen	decimal(18,2)	""
""	null	null
# Detailed Table Information	null	null

Luego se puede apreciar que es una tabla externa.

Location:	hdfs://sandbox-hdp.hortonworks.com:8020/data/raw	null
Table Type:	EXTERNAL_TABLE	null
Table Parameters:	null	null
""	EXTERNAL	TRUE

MUCHAS
GRACIAS

