



Instituto Tecnológico de Estudios Superiores de Monterrey
Campus Monterrey

“Apegándome a la Integridad Académica de los Estudiantes del Tecnológico de Monterrey, me comprometo a que mi actuación en esta actividad esté regida por la integridad académica. En congruencia con el compromiso adquirido, realizaré este trabajo de forma honesta y personal, para reflejar, a través de él, mi conocimiento y aceptar, posteriormente, la evaluación obtenida”

Inteligencia artificial avanzada para la ciencia de datos II
Gpo 502

Módulo: 1. Big Data.
Data warehouse.

Alumno:
A01750164 | Paul Martín García Morfín

Profesor:
Félix Ricardo Botello Urrutia

Fecha: 29/11/2022

Instrucciones

Utilizando un dataset con múltiples tablas, realizar lo siguiente:

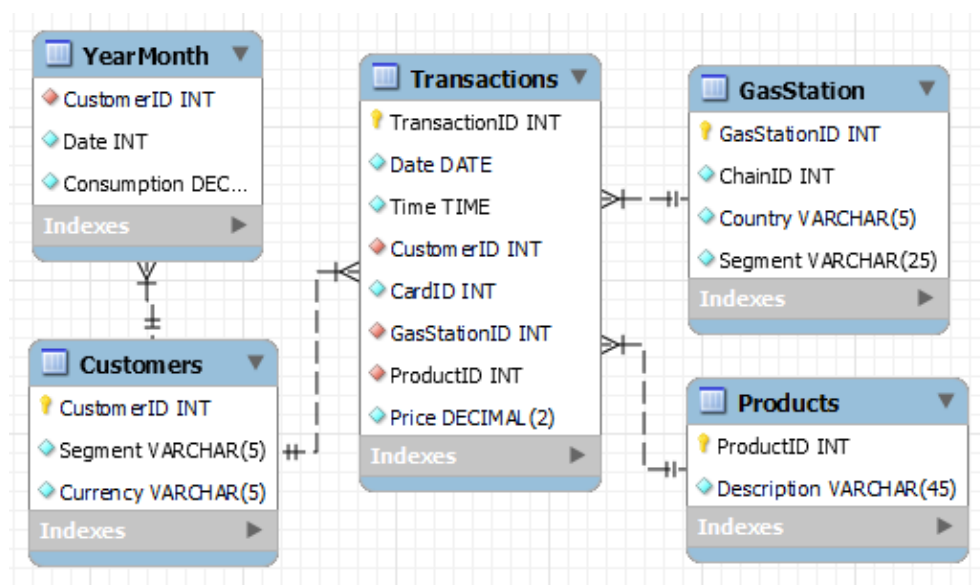
1. Realizar el proceso para crear un Data Warehouse. Se puede utilizar servicio Cloud, como AWS Redshift, Azure Synapse, Google BigQuery, etc, o localmente en sus computadoras (SQL, MySQL, etc)
2. Utilizar una herramienta de visualización (Tableau, PowerBI, etc) para mostrar información relevante para la toma de decisiones.
3. Publicar el reporte realizado.

Modelo relacional

Para esta actividad se hizo uso de la base de datos proporcionada que contiene información sobre transacciones realizadas en gasolineras, la cual contiene las siguientes tablas:

- Customers: Contiene datos sobre los clientes
- Gas station: País y segmento de las estaciones de gasolina
- Products: Listado con los tipos de combustible
- Transactions: Relaciona clientes, producto, estación de gasolina, fecha y hora y precio de la transacción.
- Year/Month: Consumos del cliente

Se utilizó SQL Workbench para la creación del esquema, en el cual se observan las relaciones y atributos de las clases mencionadas anteriormente.

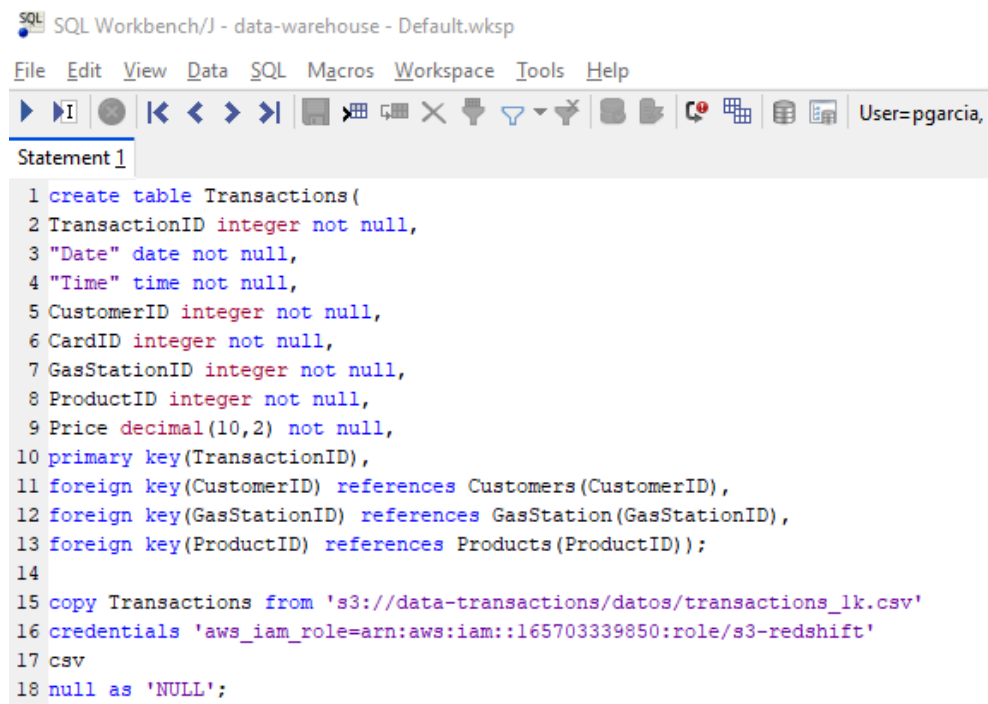


Creación del Data Warehouse

El modelo mostrado se implementó en un clúster de Amazon Redshift, siendo este el entorno de creación para el data warehouse. Para realizar este proceso fue necesario seguir los siguientes pasos:

1. Creación de un rol de IAM
2. Lanzamiento de un clúster de AWS Redshift
 - a. Configuración de la capacidad, base de datos, usuario y contraseña
3. Configuración de los grupos de seguridad para permitir la conexión
4. Uso de un bucket de AWS S3 para almacenar los archivos .csv proporcionados

Después de tener configurado todo lo anterior, se utilizó SQL Workbench/J para conectarse al clúster y así facilitar la creación de las tablas del esquema, además de la carga de datos desde los archivos .csv. Para ello, se utilizaron algunos comandos como los mostrados a continuación.



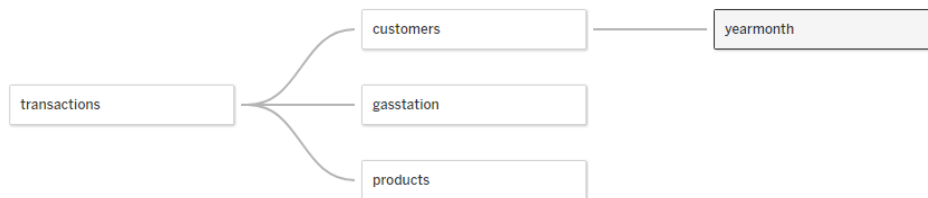
```
SQL Workbench/J - data-warehouse - Default.wksp
File Edit View Data SQL Macros Workspace Tools Help
[Navigation Icons] User=pgarcia,
Statement 1
1 create table Transactions(
2 TransactionID integer not null,
3 "Date" date not null,
4 "Time" time not null,
5 CustomerID integer not null,
6 CardID integer not null,
7 GasStationID integer not null,
8 ProductID integer not null,
9 Price decimal(10,2) not null,
10 primary key(TransactionID),
11 foreign key(CustomerID) references Customers(CustomerID),
12 foreign key(GasStationID) references GasStation(GasStationID),
13 foreign key(ProductID) references Products(ProductID));
14
15 copy Transactions from 's3://data-transactions/datos/transactions_lk.csv'
16 credentials 'aws_iam_role=arn:aws:iam::165703339850:role/s3-redshift'
17 csv
18 null as 'NULL';
```

Creación del Dashboard

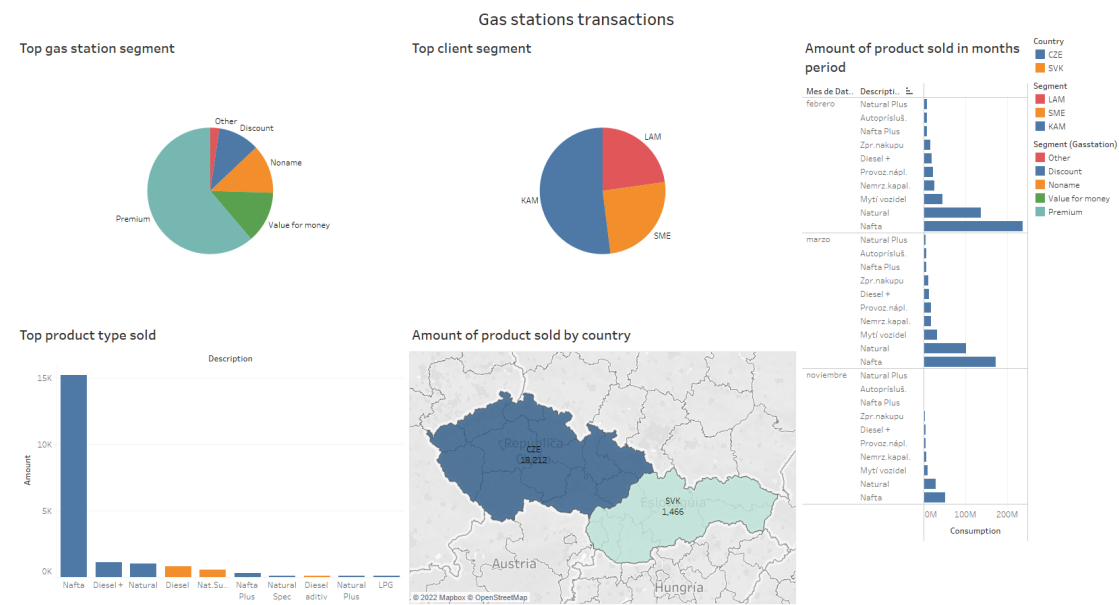
Una vez creado el data warehouse se procedió a generar los reportes de análisis. Para esta tarea se optó por utilizar Tableau, que permite conectarse a Amazon Redshift y generar las consultas de una forma más rápida y sencilla.

Lo primero fue establecer dicha conexión con el servidor de Amazon Redshift para poder acceder a los datos almacenados. Después se seleccionaron los datos con los que se iba a trabajar y se crearon las relaciones.

public



Posteriormente, con la ayuda de las distintas opciones que brinda Tableau se creó un dashboard para visualizar información de consultas de forma más gráfica y así facilitar el proceso de toma de decisiones. En el tablero se muestra información como, por ejemplo, gráficos de pastel para saber cuál segmento de clientes realiza más cantidad de transacciones así como la segmentación de las gasolineras. Otra consulta nos muestra qué tipo de gasolina es la que más se consume y en qué país, tanto en forma de gráfica de barras ordenada de forma descendente como un mapa de los países que nos permite tener una mejor visualización de los números. Por otro lado, se despliega información de ventas de los tipos de combustible según los meses, esto para que se conozca tal vez las temporadas en las cuales se vendan más ciertos productos.



Finalmente, se creó una nueva cuenta de Tableau Cloud para poder subir y compartir los tableros creados, los cuales se pueden consultar a través del siguiente enlace.

Link:

https://prod-useast-a.online.tableau.com/t/pgarcia2/views/Act3DataWarehouseC_IACD_BD/Gasstationtransactions