

Segunda entrega

Juan Manuel Ramírez, Juan Manuel Uribe y Juan Esteban Velandia

Fuente de los datos: <https://www.datos.gov.co/Econom-a-y-Finanzas/Tarjetas-de-credito-y-d-bitto/h2jg-r3zg>.

Link del repositorio GitHub: https://github.com/JuanManuel29/Proyecto_ID.git

Procedimiento para la ejecución de las consultas:

1. Ingrese al repositorio y ubíquese en la carpeta 'SQL_scripts'. Dentro de esta carpeta, encontrará el archivo 'DDL.sql' el cual contiene el código para la creación de las tablas de la base de datos y el archivo 'Insert_data.sql' el cual contiene las sentencias para traer los datos del excel a la base de datos
2. En el mismo repositorio, ubíquese en la carpeta 'Excel_files' y descargue todos los archivos
3. Ingrese al pgAdmin 4, cree la base de datos con el nombre que desee, copie y ejecute el código del DDL que encontró en el repositorio para crear las tablas
4. Ahora, cree un archivo nuevo y ejecute allí el código contenido en el archivo 'Insert_data.sql', modificando las rutas dependiendo en donde estén almacenados los archivos de excel descargados anteriormente
5. Una vez, hecho esto, ingresar al repositorio nuevamente y ubicarse en la carpeta 'Python_scripts'. Dentro de esta carpeta encontrará 2 archivos: 'credenciales.py' e 'ID_proyecto.py'. El primero contiene las credenciales necesarias para conectarse a la base de datos y el segundo contiene el script para la consulta de todos los datos de las tablas.
6. Copie sus credenciales y la conexión con la base de datos estaría lista
7. Ahora, en el archivo 'ID_proyecto.py' basta con ejecutar la función `print_all_data(table_name: str)` la cual recibe el nombre de la tabla de la cual quieren ser consultados los datos.

Nota: En la parte superior de este documento se encuentra una lista llamada 'table_names' en la cual están todos los nombres de las tablas.

8. Ejecute el script y visualice los datos en la terminal

Reglas de negocio:

1. A una subcuenta se le pueden asignar varios registros.
2. Se identifica una subcuenta por medio de un código, además posee una descripción.
3. Buscar identificar las características de las subcuentas, tipos de tarjetas e internacionalización por medio de la descripción.
4. Los registros cuentan con un número único, la fecha de registro, el número de datos por persona jurídica, persona natural y un total del número de datos que puede ser la suma de los tipos de personas o no.
5. Se debe poder identificar cuáles totales son las sumas y cuáles no, es decir, cuáles subcuentas son flujos y cuales stocks.
6. Poder identificar la cantidad de registros por subcuenta.
7. Los registros se ven agrupados por medio de las UCA. Una UCA puede tener varios registros.
8. Se debe percatar dentro de los registros, cuál fue la fecha donde más se registró cierta UCA.
9. Las UCA tienen un código único y su nombre.
10. Una clase posee un código y un nombre. Estas clasifican a las entidades por clases.
11. Las entidades requieren obligatoriamente de una clase para poder conocerse. Estas cuentan con un código que las identifica además de su nombre.
12. Las entidades ofrecen varias UCA Una entidad puede tener estar ofreciendo a varias UCA.
13. Se debe diseñar un ranking de UCA's en base a cuáles tienen más subcuentas.

Entidades, relaciones y atributos:

- Clase
 - o Código, nombre

Clase se refiere a los tipos de empresa bajo los que se agrupan los registros. Código identifica cada uno de los tipos siguiendo los valores originales de la base de datos y el nombre identifica a que tipo de entidad se hace alusión. El nombre se generó según las entidades clasificadas por cada tipo.

- Entidad

- Código, nombre

Entidad incluye a las empresas que se consideran dentro de los registros. Código identifica cada empresa dentro de su tipo, por esto es una entidad débil, nombre se refiere al nombre o razón social de la empresa.

- UCA
 - Código, nombre

UCA define los diferentes tipos de tarjetas de crédito ofertados por las entidades. Código identifica cada tipo de tarjeta de crédito y nombre se relaciona con el nombre del tipo de tarjeta.

- Subcuenta
 - Código, descripción

Subcuenta contiene los diferentes tipos de subcuentas que llevan las entidades para las UCA. Código identifica la subcuenta y tarjetas

- Registro
 - Fecha, natural, jurídica, total

Registro es una relación entre UCA y subcuenta donde se tienen las cantidades asociadas a las subcuentas para las UCA. Fecha se refiere a la fecha de corte, o final del mes, donde se genera el registro, natural y jurídica son las cantidades asociadas a las subcuentas divididas en personas (natural) e instituciones (jurídicas) y total se refiere al total de personas asociadas a las subcuentas tal que en algunos casos es la suma de los dos atributos previamente mencionados o en algunos otros casos es un acumulado hasta la fecha.

Relaciones:

- Una clase puede **clasificar** múltiples entidades.
- Entidades **ofertan** múltiples UCA.
- UCA se **ofertan** en múltiples entidades.
- UCA **agrupa** varios registros.
- A una subcuenta se le **asignan** varios registros.

Diagrama entidad relación:

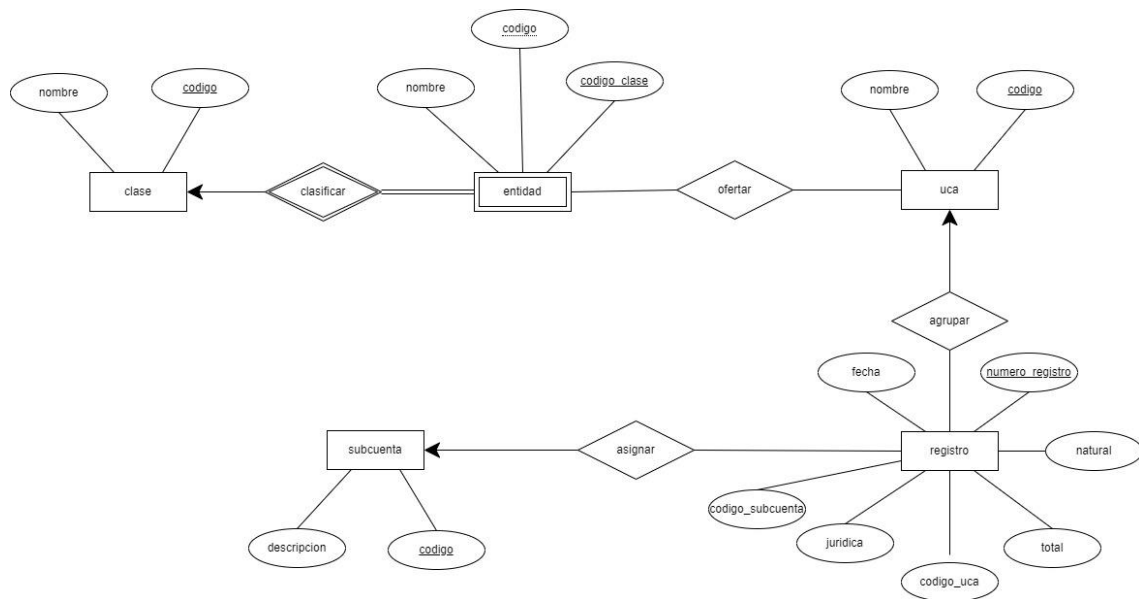
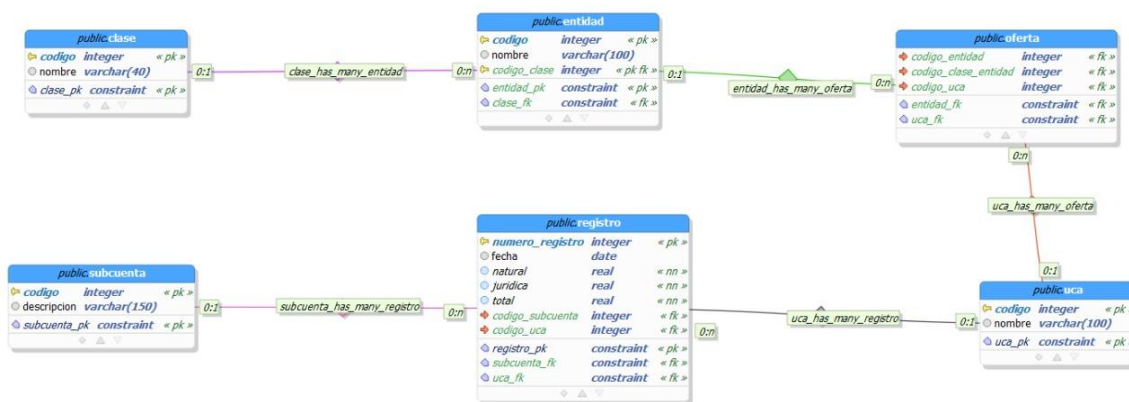


Diagrama relacional (normalizado en tercera forma normal):



Posibles casos de análisis:

- Identificar cuales subcuentas son flujos ($\text{natural} + \text{juridica} = \text{total}$) y cuales stocks ($\text{natural} + \text{juridica} \neq \text{total}$).
- Hallar distribución de tipos de empresas que ofrecen tarjetas crediticias
- Identificar las subcuentas menos utilizadas (cantidad de registro)
- Comparar compras y avances realizados con tarjetas (nacional vs internacional, crédito vs débito) por UCA

Procedimiento para la ejecución de las consultas:

1. Ingrese al repositorio y ubíquese en la carpeta 'SQL_scripts'. Dentro de esta carpeta, encontrará el archivo 'DDL.sql' el cual contiene el código para la creación de las tablas de la base de datos y el archivo 'Insert_data.sql' el cual contiene las sentencias para traer los datos del excel a la base de datos
2. En el mismo repositorio, ubíquese en la carpeta 'Excel_files', descargue todos los archivos y colóquelos en la carpeta C\:
3. Ingrese al pgAdmin 4, cree la base de datos con el nombre que desee, copie y ejecute el código del DDL que encontró en el repositorio para crear las tablas
4. Ahora, cree un archivo nuevo y ejecute allí el código contenido en el archivo 'Insert_data.sql', modificando las rutas dependiendo en donde estén almacenados los archivos de excel descargados anteriormente
5. Una vez, hecho esto, ingresar al repositorio nuevamente y ubicarse en la carpeta 'Python_scripts'. Dentro de esta carpeta encontrará 2 archivos: 'credenciales.py' e 'ID_proyecto.py'. El primero contiene las credenciales necesarias para conectarse a la base de datos y el segundo contiene el script para la consulta de todos los datos de las tablas.
6. Copie sus credenciales y la conexión con la base de datos estaría lista.
7. Ahora, en el archivo 'ID_proyecto.py' basta con ejecutar la función *print_all_data(table_name: str)* la cual recibe el nombre de la tabla de la cual quieren ser consultados los datos.

Nota: En la parte superior de este documento se encuentra una lista llamada 'table_names' en la cual están todos los nombres de las tablas.

8. Ejecute el script y visualice los datos en la terminal.