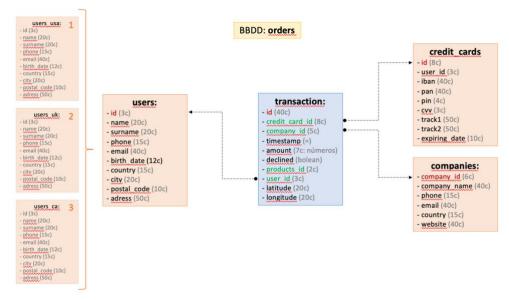
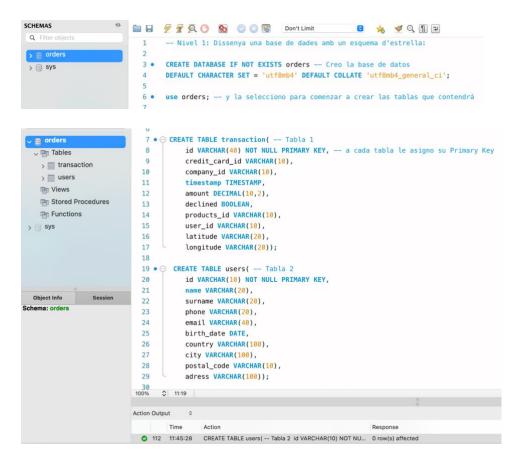
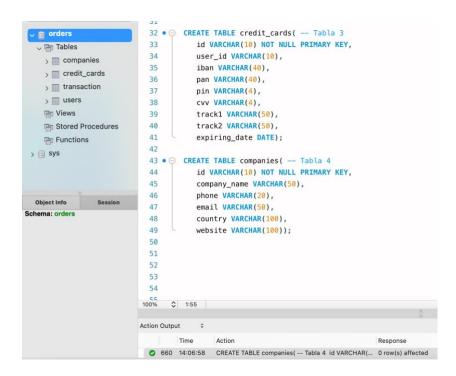
Descàrrega els arxius CSV, estudia'ls i dissenya una base de dades amb un esquema d'estrella que contingui, almenys 4 taules de les quals puguis realitzar les següents consultes:



Creamos la BBDD y las tablas:





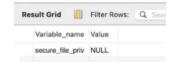
En la tabla 'transaction' elegí la mayoría de datos en formato VARCHAR (con una amplitud dependiente de los caracteres que tuviera nuestro campo) porque la mayoría de campos contenían números y letras o caracteres. En el caso del campo 'timestamp' como su nombre indica su tipo de dato ha de ser TIMESTAMP, 'amount' es decimal porque se trata de dinero y declined es un boleano porque solo acepta valores de 0 o 1.

En la tabla 'users' todos los campos son VARCHAR por lo mismo que en la anterior, excepto el campo 'birth_date' que como su nombre indica es una fecha, ya que así será más fácil de trabajar y manipular en el futuro. Nombre, apellido y país los dejo como VARCHAR en lugar de texto por si existiera alguno de ellos que contuviera un carácter especial no contemplado en otros países (por ejemplo la 'ñ' de 'España' que no existe en los teclados de países de habla inglesa).

En el caso de la tabla 'credit_card', al igual que en 'users' he definido todos los campos como VARCHAR exceptuando el campo 'expiring_date' que es una fecha. En este caso tanto 'cvv' como 'pin' podrían ser considerados INT (enteros) pero entonces tendríamos problemas si existiera algún código que fuera, por ejemplo: 0001 porque lo detectaría como 1. El IBAN (International Bank Account Number), utilizado para identificar de forma estandarizada cualquier cuenta sea del país que sea, es una serie alfanumérica formada por un máximo de 34 caracteres dependiendo del país, por eso lo he definido como VARCHAR (40) para darle algo de margen. El CVV o código de verificación de la tarjeta es un código de tres o cuatro números utilizado para verificar que el usuario de la compra es el poseedor de la tarjeta por eso lo he definido como VARCHAR (4) porque es el máximo número de caracteres que puede aceptar. El PIN es el código de seguridad de la tarjeta que he definido como VARCHAR (4) porque suele ser de 4 números.

Por último, en la tabla 'companies' todos los datos son de tipo VARCHAR por lo anteriormente explicado.

 Al intentar cargar los csv para añadirlos a las tablas me daba error ya que la exportación e importación de datos estaba desactivada porque las versiones de MySQL anteriores a la 8.0 no tienen un archivo llamado my.cnf, y por eso al ejecutar: SHOW VARIABLES LIKE "secure_file_priv" Nos devuelve el valor NULL.

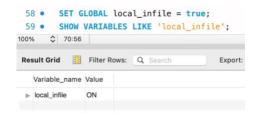


Para solucionar esto:

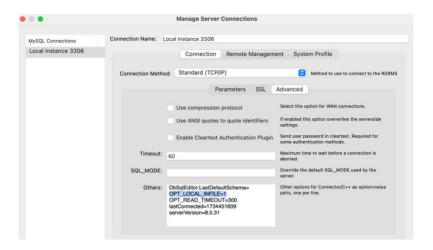
1º Creé el archivo /etc/my.cnf que no existía y le añadí los siguientes valores:



2º Me aseguré que la variable 'local infile' estuviera activa igualando su valor a true:



3º Como continuaba dándome error, edité la conexión en la subpestaña 'Advanced' de la casilla 'Others' añadiendo: OPT_LOCAL_INFILE=1



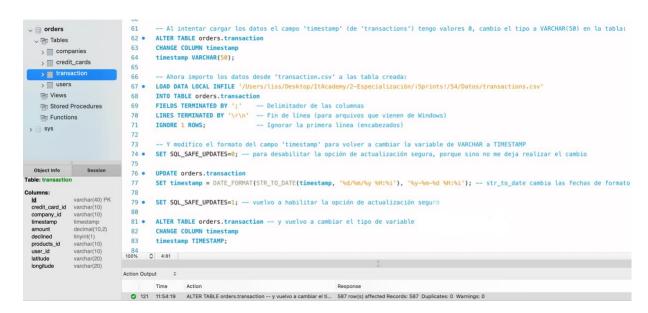
Y entonces funcionó:



Ahora sí, cargo los csv para añadirlos a las tablas:

Tabla 1: 'transaction':

Al intentar cargar los datos el campo 'timestamp' me salen valores a 0, por eso cambio el campo de TIMESTAMP a VARCHAR(50) en la tabla creada por mí, importo los datos de los archivos csv y modifico el formato de los datos del campo para volver a cambiar la variable de VARCHAR a TIMESTAMP.



Me sale un WARNING al importar 'transaction':



Pero al visualizar la tabla todo está correcto y como menciona la columna 'products_id' deduzco que es porque esta tiene varios valores separados por comas en algunas de sus filas.

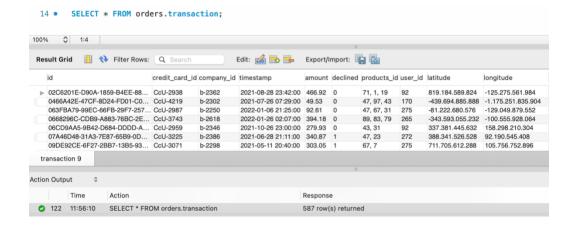
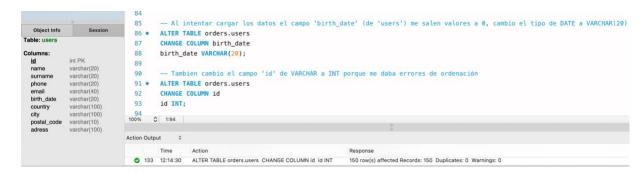


Tabla 2: 'users':

Al intentar cargar los datos el campo 'birth_date' me salen valores a 0, por eso cambio el campo de DATE a VARCHAR(20) en la tabla creada por mí. También modifico el campo 'id' de VARCHAR a INT porque me daba errores de ordenación al intentar cargar los datos.



Importo los datos de los archivos csv:

```
→ 

⊜ orders

√ Tables

                                         -- Luego importo los datos desde los diferentes archivos de csv (users_usa, users_uk y users_ca) a la tabla creada:
    > m companies
                                98 •
                                        LOAD DATA LOCAL INFILE '/Users/liss/Desktop/ItAcademy/2-Especialización/iSprints!/S4/Datos/users_usa.csv
    > m credit_cards
                                99
                                        INTO TABLE orders.users
    > m transaction
                                100
                                        FIELDS TERMINATED BY '.'
                                                                        -- Delimitador de las columnas
                                        OPTIONALLY ENCLOSED BY "" -- Delimitador de texto (comillas dobles), para que me coja bien la fecha y la dirección
                                101
                                        LINES TERMINATED BY '\r\n' -- Fin de línea (para archivos que vienen de Windows)
                                102
                                        IGNORE 1 LINES;
                                103
                                                                        -- Ignorar la primera línea (encabezados)
    R Stored Procedures
                                104
    ⊞ Functions
                                        LOAD DATA LOCAL INFILE '/Users/liss/Desktop/ItAcademy/2-Especialización/iSprints!/S4/Datos/users_uk.csv'
 > 🗎 sys
                                        INTO TABLE orders.users
                                107
                                        FIELDS TERMINATED BY 1.1
                                                                        -- Delimitador de las columnas
                                        OPTIONALLY ENCLOSED BY '"' — Delimitador de texto (comillas dobles), para que me coja bien la fecha y la dirección
LINES TERMINATED BY '\r\n' — Fin de línea (para archivos que vienen de Windows)
                                108
                                109
  Object Info
                               110
                                        IGNORE 1 LINES:
                                                                       -- Ignorar la primera línea (encabezados)
Table: users
                               111
                               112 • LOAD DATA LOCAL INFILE '/Users/liss/Desktop/ItAcademy/2-Especialización/iSprints!/S4/Datos/users ca.csv
Columns:
             int PK
                                        INTO TABLE orders.users
                               113
                                        FIELDS TERMINATED BY ','
                               114
                                                                        -- Delimitador de las columnas
                                        OPTIONALLY ENCLOSED BY '"' -- Delimitador de texto (comillas dobles), para que me coja bien la fecha y la dirección

LINES TERMINATED BY '\r\n' -- Fin de línea (para archivos que vienen de Windows)
                               115
  email
 birth_date
                               117
                                        IGNORE 1 LINES;
                                                                        -- Ignorar la primera línea (encabezados)
 country
                                     ☼ 1:96
 postal_code
                                        Time
                                                 Action
                                                                                                  Response
                                 2 136 12:17:43 LOAD DATA LOCAL INFILE '/Users/liss/Desktop/ItAcademy/2... 75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0
```

Modifico el formato de los datos del campo para volver a cambiar la variable de VARCHAR a DATE y visualizo la tabla:

```
Object Info

Objec
```

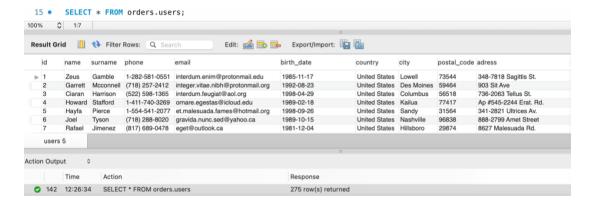
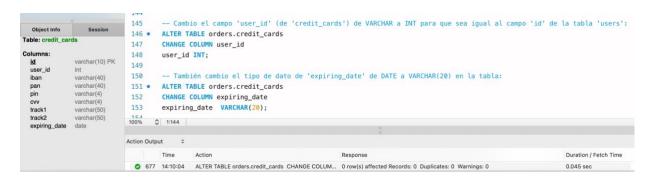


Tabla 3: 'credit cards':

Primero cambio el campo 'user_id' a INT para que coincida con el campo 'id' de la tabla 'users' y el campo 'expiring_date' a VARCHAR para que acepte los datos y luego lo volveré a cambiar a DATE:



Importo los datos del archivo csv, cambio el formato de fecha de 'expiring_date', vuelvo a cambiar el tipo de variable a DATE y visualizo mi tabla:

```
-- Importo los datos desde 'credit_cards.csv' a la tabla creada:
 145 • LOAD DATA LOCAL INFILE '/Users/liss/Desktop/ItAcademy/2-Especialización/iSprints!/S4/Datos/credit_cards.csv'
 146
        INTO TABLE orders.credit_cards
 147
        FIELDS TERMINATED BY ','
                               -- Delimitador de las columnas
 148
        LINES TERMINATED BY '\n'
                                -- Fin de línea
                                 -- Ignorar la primera línea (encabezados)
 150
 151
        -- Y modifico el formato del campo 'expiring_date' para volver a cambiar la variable de VARCHAR a DATE
 152 • SET SQL_SAFE_UPDATES=0; -- para desabilitar la opción de actualización segura
 153
        UPDATE orders.credit_cards
        SET expiring_date = DATE_FORMAT(STR_TO_DATE(expiring_date, '%m/%d/%Y'), '%Y-%m-%d') -- str_to_date cambia las fechas de formato
 155
 156
        WHERE credit_cards.id IS NOT null;
 157
 158 • SET SOL SAFE UPDATES=1: -- vuelvo a habilitar la opción de actualización segura
 159
 160 • ALTER TABLE orders.credit cards -- v vuelvo a cambiar el tipo de variable de 'expioring date' a DATE
 161
        CHANGE COLUMN expiring_date
 162
        expiring_date DATE;
      C 1:140
Action Output
        Time
               Action
                                                      Response
```

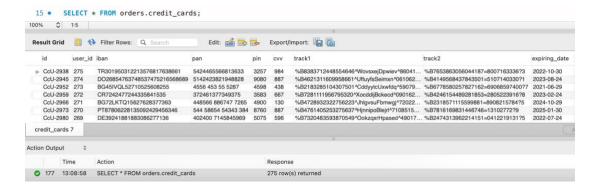
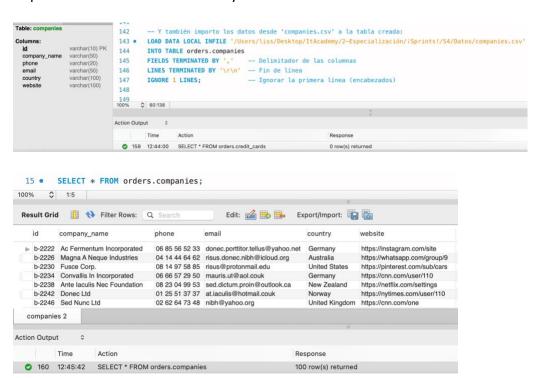


Tabla 4: 'companies:

Importo los datos del archivo csv y visualizo mi tabla:



Añado las Foreign Keys que relacionaran las tablas entre ellas en este esquema de estrella:

```
167
                                               -- Ahora añado las Foreign Keys que relacionaran las tablas entre ellas en este esquema de estrella
                                             ALTER TABLE orders transaction
                                    168 •
                                    169
                                             ADD FOREIGN KEY (credit_card_id) REFERENCES credit_cards(id);
                                    170
                                    171 •
                                            ALTER TABLE orders.transaction
                                             ADD FOREIGN KEY (company_id) REFERENCES companies(id);
                                    172
Columns:
                                    174 •
                                             ALTER TABLE orders.transaction
 olumns:

id

credit_card_id

company_id

timestamp

amount

declined

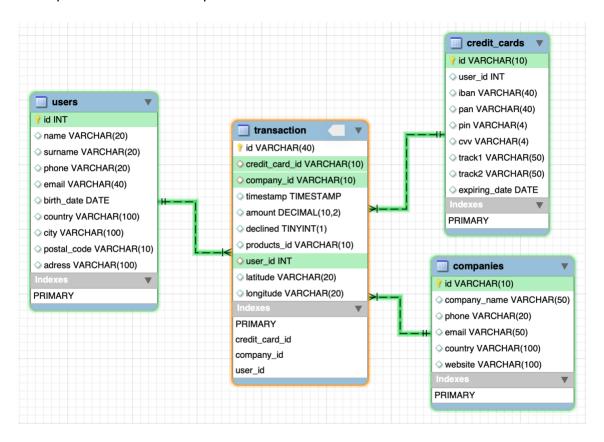
products_id

user_id

latitude

longitude
                  varchar(40) Pivarchar(10) varchar(10) timestamp decimal(10,2) tinyint(1) varchar(10) int
                                    175
                                             CHANGE COLUMN user id
                                    176
                                             user_id INT; -- cambio el campo 'user_id' a INT para que sea igual al campo 'id' de la tabla 'users'
                                    177
                                    178 . ALTER TABLE orders transaction
                                             ADD FOREIGN KEY (user_id) REFERENCES users(id);
                                    179
                                    100%
                                  2 182 17:14:47 ALTER TABLE orders.transaction ADD FOREIGN KEY (user_i... 587 row(s) affected Records: 587 Duplicates: 0 Warnings: 0
```

Y así queda mi BBDD con esquema de estrella:

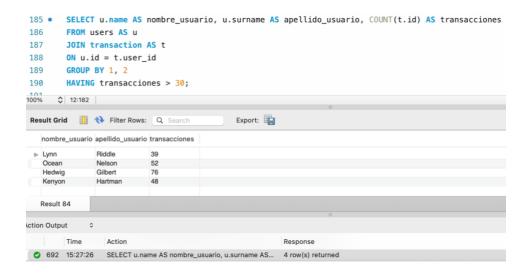


En el diagrama observamos que la tabla de hechos sería 'transaction' y las tablas de dimensiones 'companies', 'users' y 'credit_cards'. La relación entre las tres tablas se establece a través de las Primary Keys de las tablas de dimensiones (id de 'companies', id de 'users' e id de 'credit_cards') y las Foreign Keys de la tabla de hechos: company_id que se relaciona con id de 'company', user_id que se relaciona con la id de 'users' y credit_card_id que se relaciona con id de 'credit_card'.

^{*}Aunque también había un archivo de csv que contenía datos de una tabla llamada 'products' por ahora no lo añado a mi base de datos porque no lo usaré hasta el nivel 3.

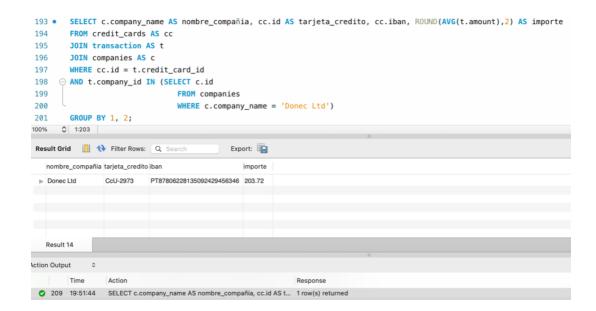
Exercici 1

Realitza una subconsulta que mostri tots els usuaris amb més de 30 transaccions utilitzant almenys 2 taules.



Exercici 2

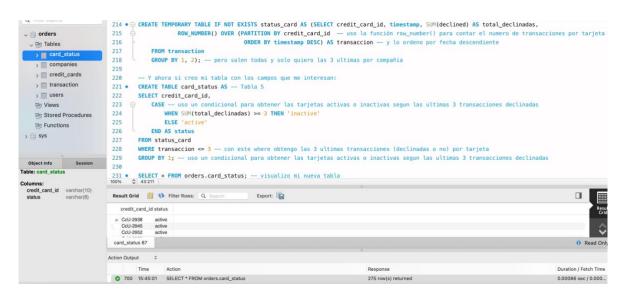
Mostra la mitjana d'amount per IBAN de les targetes de crèdit a la companyia Donec Ltd, utilitza almenys 2 taules.



Crea una nova taula que reflecteixi l'estat de les targetes de crèdit basat en si les últimes tres transaccions van ser declinades i genera la següent consulta:

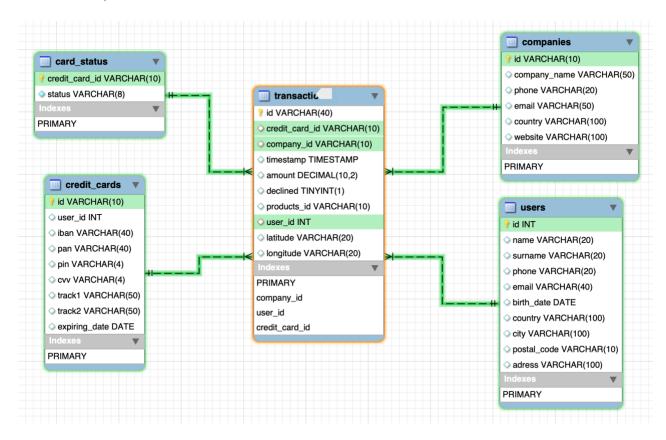
```
206 • SELECT credit_card_id, SUM(declined) — compruebo cuantas tarjetas han sido declinadas más de una vez en general
 207
        FROM transaction
        GROUP BY 1;
 208
 209
        -- Ninguna tarjeta ha sido declinada más de una vez
710
100% $ 19:204
 ▶ CcU-2938
   CcU-2945
CcU-2952
   CcU-2959
   CcU-2966
   CcU-2973
   Result 23
Action Output 0
      Time
               Action
                             Response
250 13:26:11 SELECT credit_c... 275 row(s) returned
```

 Primero creo una tabla temporal, luego creo la tabla con los campos que me interesan y la visualizo:



Por último, añado su PK y la relaciono con una FK de la tabla 'transaction':

- Y así queda nuestra BBDD ahora:



La nueva tabla 'card_status' nos queda relacionada con la tabla de hechos 'transaction' a través de su PK credit_card_id con la FK (del mismo nombre) de la tabla 'transaction', de manera que si se actualizan los datos de las transacciones se actualice el estado de las tarjetas activas o inactivas.

Exercici 1

Quantes targetes estan actives?

- Todas las tarjetas están activas:

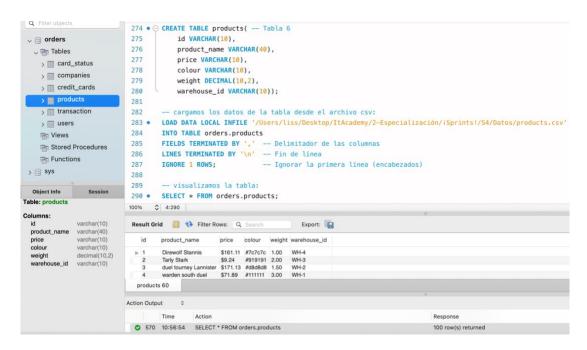


- Lo comprobamos mirando que no haya tarjetas inactivas:

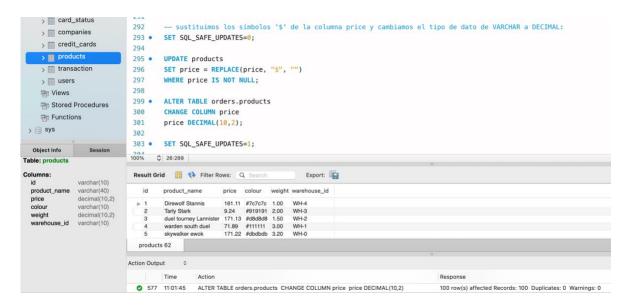


Crea una taula amb la qual puguem unir les dades del nou arxiu products.csv amb la base de dades creada, tenint en compte que des de transaction tens product ids. Genera la següent consulta:

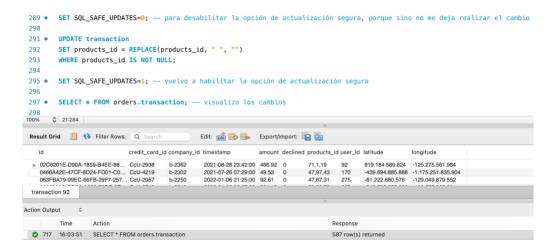
- Primero creamos la tabla 'products' y le añadimos los datos del csv con el mismo nombre:



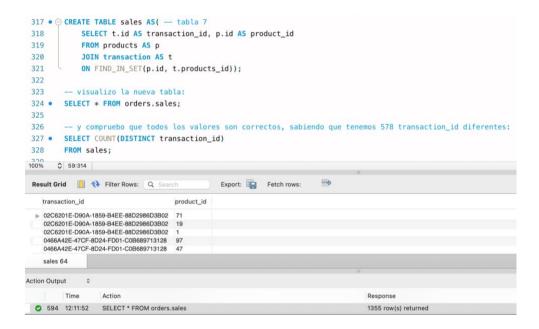
 Elimino el símbolo de dólar de la columna 'price' de 'products' para convertirla en decimal y trabajar mejor con sus datos ya que se trata de dinero:



 También elimino los espacios después de las comas de 'products_id' de la tabla 'transaction', para que al compararla con 'id' de la tabla 'products' no nos tenga en cuenta los espacios vacíos:

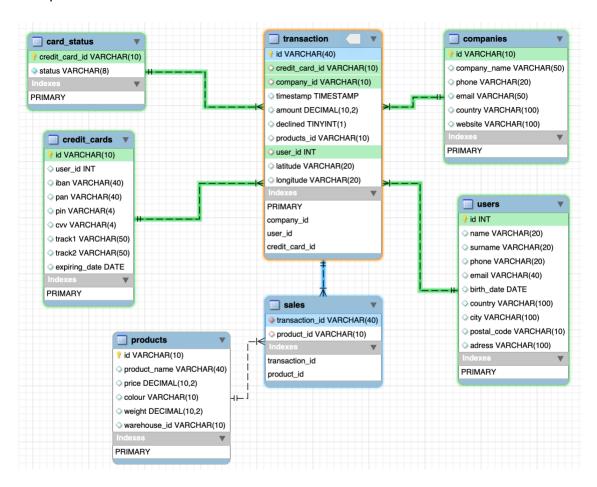


 Y creo la tabla 'sales' que es una tabla puente que me relaciona las tablas 'products' y 'transaction':



 Por último, reo la PK de la tabla 'products' y le añado las FKs a la tabla 'sales' que la relacionan con 'transaction' y 'products':

- Y así queda nuestra BBDD final:



La tabla puente 'sales' nos queda relacionada con la tabla de hechos 'transaction' a través de una FK llamada transaction_id que se relaciona con la PK (id) de la tabla 'transaction' y también se relaciona a través de la FK producto_id con la PK (id) de la tabla 'products' que es una nueva tabla de dimensiones que añadimos a nuestra BBDD.

Exercici 1

Necessitem conèixer el nombre de vegades que s'ha venut cada producte.

