Tasca S4.01. Creació de Base de Dades

Nivel 1

Descarga los archivos CSV, estúdialas y diseña una base de datos con un esquema de estrella que contenga, al menos 4 tablas de las que puedas realizar las siguientes consultas:

Creamos primero la base de datos transacciones donde estarán todas las tablas que cargaremos desde los archivos CSV's.

Mediante use transacciones; pondremos por defecto esta base de datos.

Descargamos en nuestro ordenador los archivos CSV's y los visualizamos con Visual Studio Code.



Creamos las tablas con Create table y verificamos si existen con el "IF NOT EXISTS".

También creamos las Primary key en cada tabla y en las tablas products y users crearemos auto_increment, y tipo de dato INT

En la mayoría de caso creamos el tipo de dato VARCHAR (255), y en casos que observemos que son enteros: INT, en Price y amount: decimal (10,2), ya que manejaran datos de precios y montos, y en la columna timestamp tipo de dato TIMESTAMP, en las columnas de lat y longitude: FLOAT, ya que son coordenadas.

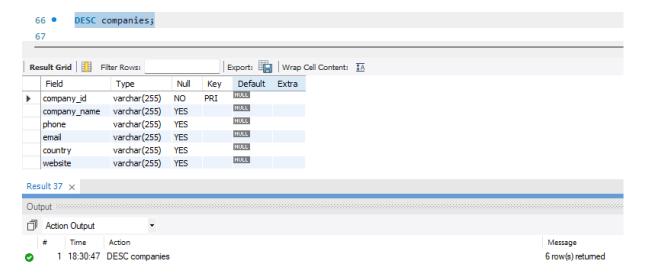
```
CREATE DATABASE IF NOT EXISTS transacciones;
   4 •
   5 •
         USE transacciones;
          -- creamos las tablas
   8 ● ⊝ CREATE TABLE IF NOT EXISTS companies (
   9
             company_id VARCHAR(255) PRIMARY KEY,
              company_name VARCHAR(255),
  10
  11
              phone VARCHAR(255),
              email VARCHAR(255),
  12
              country VARCHAR(255),
  13
  14
              website VARCHAR(255)
  15
  17 ● ⊝ CREATE TABLE IF NOT EXISTS credit_cards (
             id VARCHAR(255) PRIMARY KEY,
  18
  19
             user_id INT,
  20
             iban VARCHAR(255).
             pan VARCHAR(255),
             pin VARCHAR(255),
  22
             CVV INT,
  23
  24
             track1 VARCHAR(255),
             track2 VARCHAR(255).
  25
              expiring_date VARCHAR(255)
  27
  28
  29 ullet CREATE TABLE IF NOT EXISTS products (
             id INT PRIMARY KEY AUTO_INCREMENT,
  30
  31
              product_name VARCHAR(255),
  32
             price DECIMAL(10,2),
  33
             colour VARCHAR(255),
  34
             weight VARCHAR(255),
              warehouse_id VARCHAR(255)
  35
  37
Output
Action Output
       Time
                 Action
                                                                                                               Message
4 19:47:41 CREATE DATABASE IF NOT EXISTS transacciones
                                                                                                              1 row(s) affected
      5 19:47:41 USE transacciones
                                                                                                              0 row(s) affected
6 19:47:41 CREATE TABLE IF NOT EXISTS companies (company_id VARCHAR(255) PRIMARY KEY, company_name VA... 0 row(s) affected
      7 19:47:41 CREATE TABLE IF NOT EXISTS credit_cards (id VARCHAR(255) PRIMARY KEY, user_id INT, iban VARCH... 0 row(s) affected
```

8 19:47:41 CREATE TABLE IF NOT EXISTS products (id INT PRIMARY KEY AUTO_INCREMENT, product_name VARCH... 0 row(s) affected

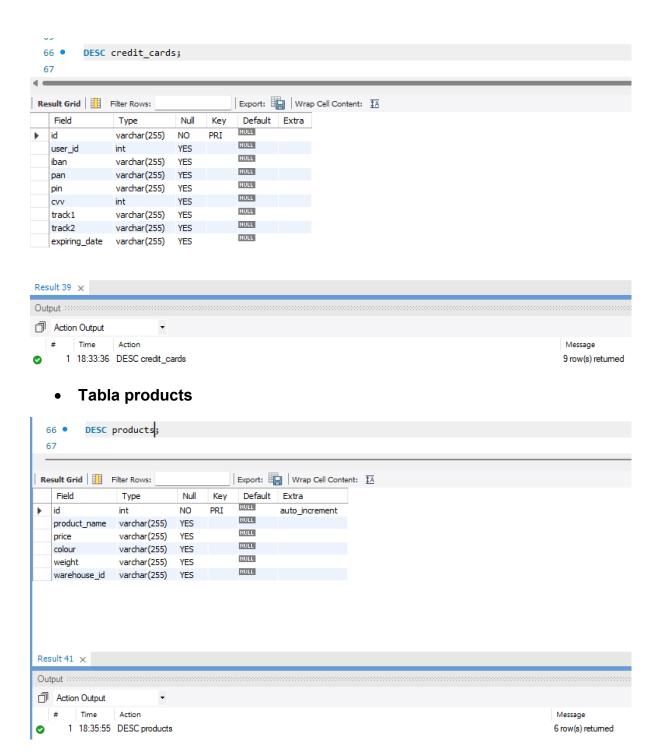
```
37
  38 ● ⊖ CREATE TABLE IF NOT EXISTS transactions (
            id VARCHAR(255) PRIMARY KEY,
  39
  40
             card id VARCHAR(255).
  41
             business_id VARCHAR(255),
             timestamp TIMESTAMP,
             amount DECIMAL(10, 2),
  43
  44
             declined TINYINT,
             product_ids INT,
              user_id INT,
  47
             lat FLOAT.
  48
              longitude FLOAT
  50
  51
           -- los campos de las tablas user_ca, user_uk, users_usa son iguales en estructura, asi que uniremos los 3 archivos csv en una sola tabla users
  53 ● ⊖ CREATE TABLE IF NOT EXISTS users (
  54
             id INT PRIMARY KEY AUTO_INCREMENT,
  55
              name VARCHAR(255),
             surname VARCHAR(255),
             phone VARCHAR(255),
  57
  58
              email VARCHAR(255),
            birth_date VARCHAR(255),
             country VARCHAR(255),
  60
  61
             city VARCHAR(255),
             postal_code VARCHAR(255),
              address VARCHAR(255)
  63
  64
          -- - Ejercicio 1
  67
          -- Realiza una subconsulta que muestre a todos los usuarios con más de 30 transacciones utilizando al menos 2 tablas.
Output
Action Output
   1 20:11:18 CREATE TABLE IF NOT EXISTS transactions ( id VARCHAR(255) PRIMARY KEY, card_id VARCHAR(255), b... 0 row(s) affected
2 20:11:18 CREATE TABLE IF NOT EXISTS users (id INT PRIMARY KEY AUTO_INCREMENT, name VARCHAR(255), sumame... 0 row(s) affected
```

Comprobamos que se hayan creado las tablas con describe:

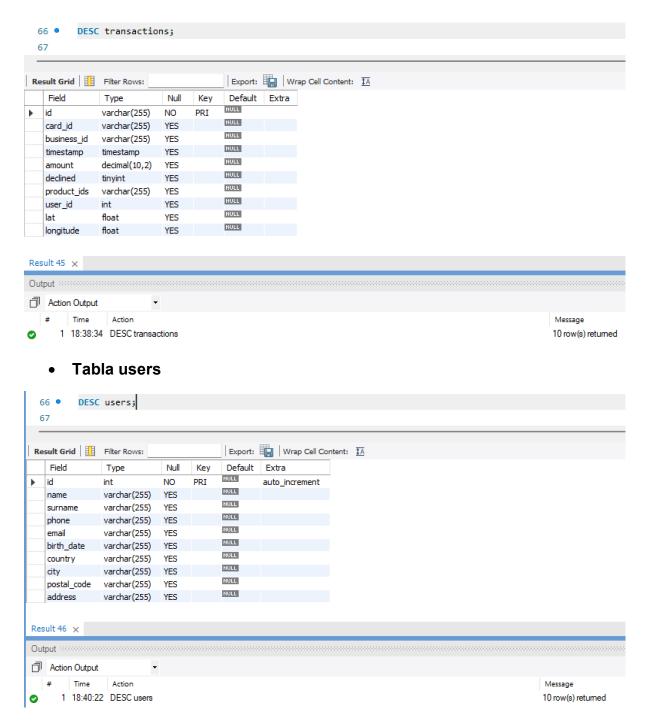
Tabla company



• Tabla credit_cards



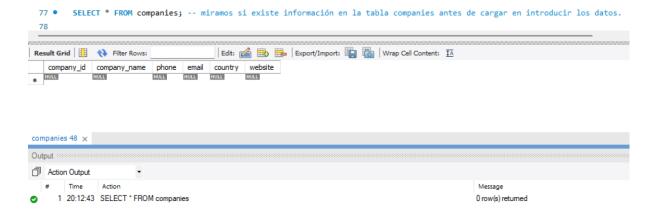
• Tabla transactions



Introducción de datos en las tablas creadas

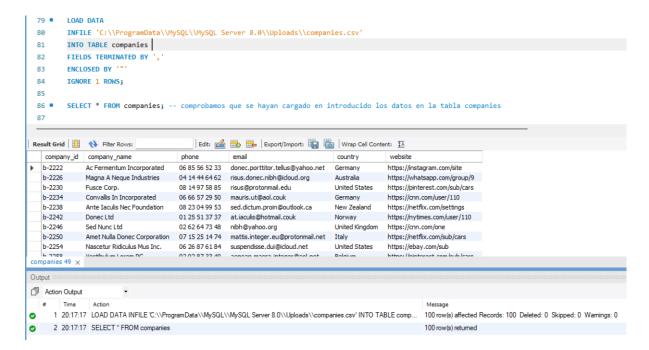
- Tabla companies

Previamente a la introducción de datos en las tablas, nos cercioramos de que no haya datos en las tablas, adjunto imagen abajo de una tabla companies, por ejemplo.



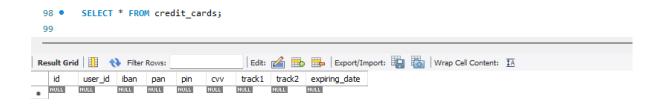
Después procederemos a la introducción de los datos:

Usamos Load data infile para la carga del archivo csv, se pone la ruta del archivo, especificamos que los campos estén separados por comas, y que tome los datos encerrados por "" como un campo, y que ignore la primera línea. Y finalmente, comprobamos nuevamente que hayan cargado todos los datos en la tabla.



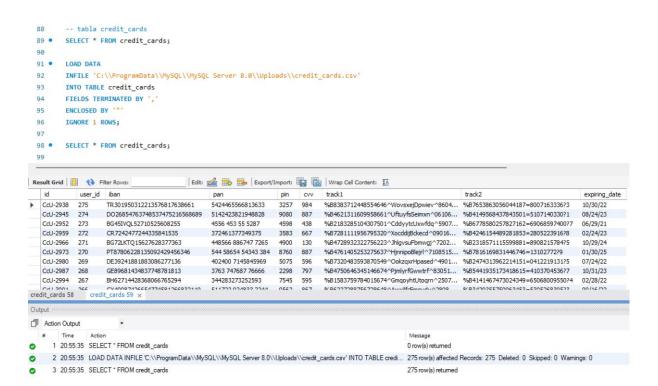
- Tabla credit_cards

Comprobamos que no exista datos en la tabla credit cards.





Cargamos los datos del archivo csv correspondiente a credit_cards e introducimos los datos, especificamos que separe los campos terminados en comas y que todo lo que este entre comillas lo considere un campo, e ignore la primera fila, al tener los encabezados ya creados en la creación de la tabla vacía. Verificamos nuevamente que hayan cargado todos los datos en la tabla.



- Tabla products

Hacemos lo mismo que para los anteriores, verificamos que no tenga datos.

Cargamos el archivo csv de products, especificamos que este separado por comas cada campo, que incluya lo que está encerrado entre comillas como un campo, y que ignore la primera fila.

Verificamos nuevamente que hayan cargado todos los datos en la tabla.

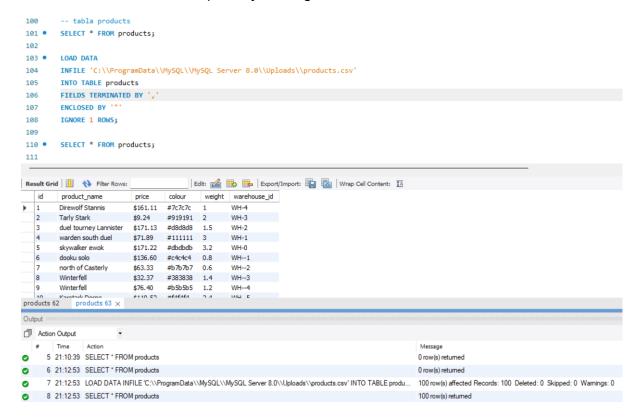


Tabla transactions

Hacemos en este caso un cambio en las especificaciones de la introducción, ya que los datos están separados por punto y coma (;) entonces lo cambiamos campos terminados en ";".

Finalmente comprobamos que los datos se hayan cargado e introducido en la tabla transactions .

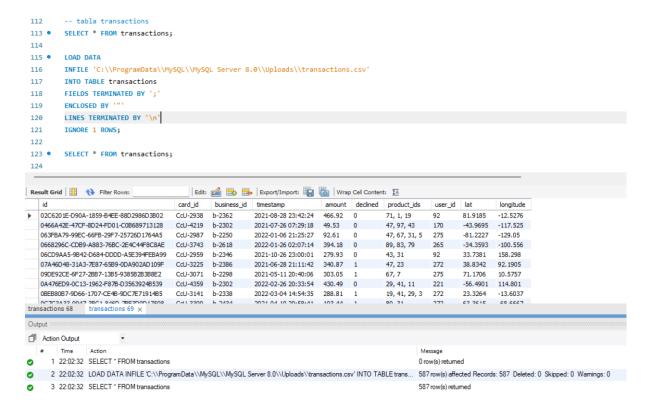
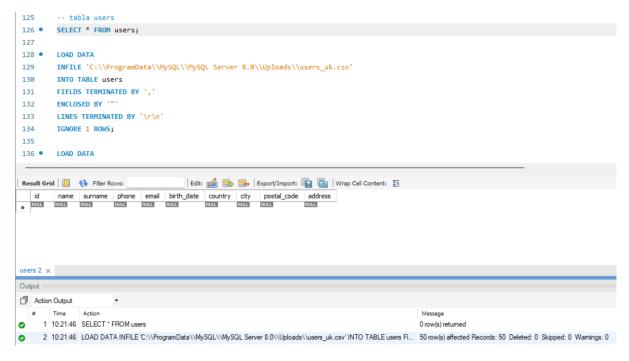


Tabla users

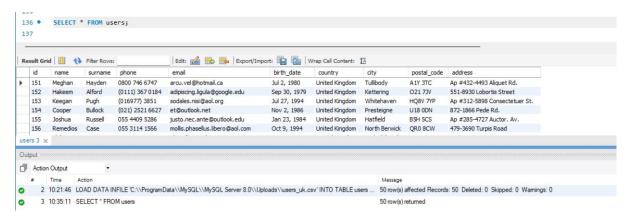
Para montar esta tabla, hay 3 archivos csv que tienen la misma estructura de campos, es por eso por lo que adjuntamos una a una cada archivo a la tabla para formar una sola llamada users.

Los datos están separados por comas, y en este caso especificamos:

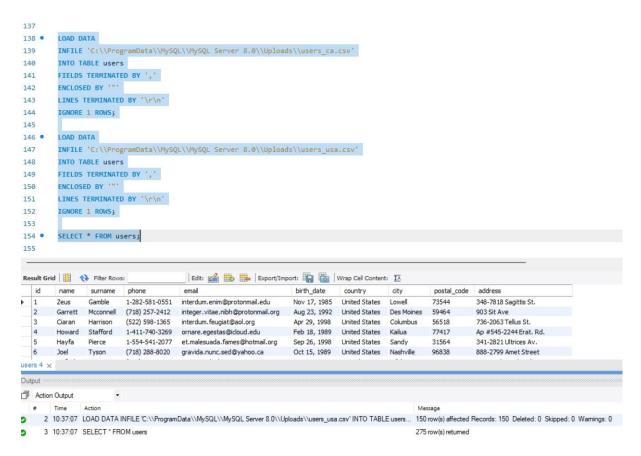
"LINES TERMINATED BY '\r\n'", este permite leer archivos csv que provienen de Windows, solo hemos usado este formato para estos archivos de users, en el resto no ha hecho falta.



Se comprueba que se hayan ingresado los datos:

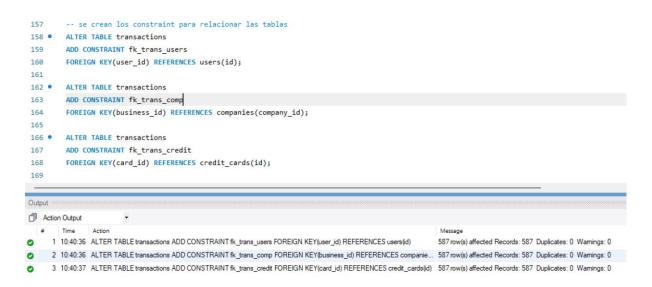


Luego continuamos ingresando los datos de los archivos csv restantes:



CREAR CONSTRAINT

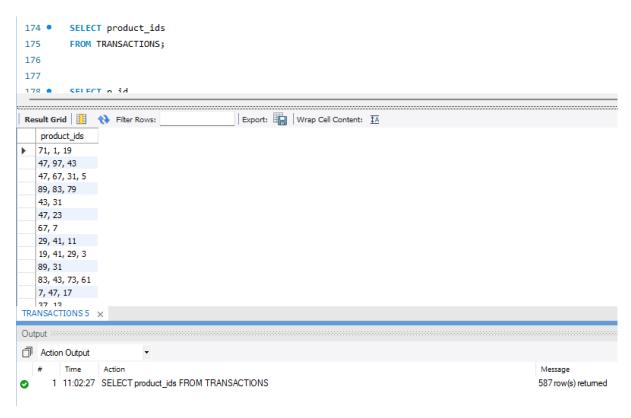
Para relacionar las tablas crearemos las constraint de la foreign key, relacionaremos las tablas users, companies y credit cards con la tabla transactions.



Para relacionar la tabla productos con transactions, nos da un error:

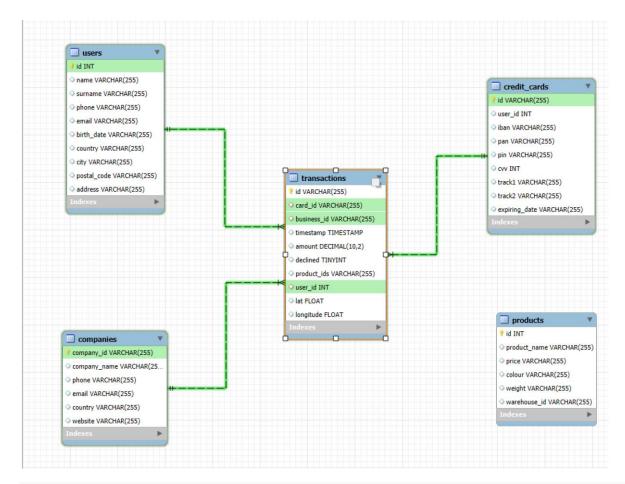
'Error Code: 3780. Referencing column 'product_ids' and referenced column 'id' in foreign key constraint 'fk_trans_prod' are incompatible'.

Este error se refiere a que las columnas que queremos relacionar en ambas tablas son incompatibles y se refiere a que en la tabla transactions la columna product_ids contiene varios valores en cada celda, que son los productos comprados en esa transaccion.



Para nuestro proyecto, dejaremos momentáneamente sin relacionar la tabla products, y quedaría nuestro diagrama de la siguiente forma:

Sería un modelo en estrella donde la tabla de hechos es transactions y las tablas dimensiones serían: users, credit_cards, companies.



- Exercici 1

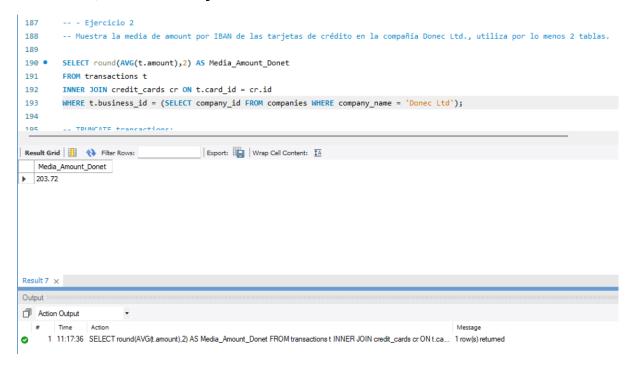
Realitza una subconsulta que mostri tots els usuaris amb més de 30 transaccions utilitzant almenys 2 taules.

```
177
       -- - Ejercicio 1
178
       -- Realiza una subconsulta que muestre a todos los usuarios con más de 30 transacciones utilizando al menos 2 tablas.
179 • SELECT u.id, COUNT(*) AS Num_Transacciones
180
      FROM transactions t
     INNER JOIN users u ON t.user_id = u.id
181
     GROUP BY u.id
183
      HAVING Num_Transacciones > 30
184
       ORDER BY Num_Transacciones DESC;
Export: Wrap Cell Content: IA
  id
       Num_Transacciones
 272
       76
  267 52
  275
       48
 92 39
```



- Exercici 2

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



Nivell 2

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:

```
201
        -- Nivell 2
        -- 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:
203
204 • CREATE TABLE activ_card AS
205
        SELECT card_id,
206 ⊝
                 CASE
207
                     WHEN sum(declined) >= 3 THEN ' no activo'
208
                    ELSE 'activo
209
                 END AS actividad
210

→ FROM (SELECT card_id, declined,
211
                ROW NUMBER() OVER (PARTITION BY card id ORDER BY timestamp DESC) AS row num
212
        FROM transactions
213
        ) AS hist_card
        WHERE row_num <=3
214
215
        GROUP BY card_id;
217 • SELECT * FROM activ card;
218
                                         Export: Wrap Cell Content: TA
card_id
            actividad
▶ CcU-2938
            activo
   CcU-2945 activo
   CcU-2952 activo
   CcU-2959 activo
   CcU-2966
           activo
   CcU-2973 activo
   CcU-2980 activo
   CcU-2987 activo
   CcU-2994 activo
   CcU-3001 activo
   CcU-3008 activo
Output ::::
Action Output
                                                                                                   Message
1 12:08:30 CREATE TABLE activ_card AS SELECT card_id, CASE WHEN sum(declined) >= 3 THEN 'no activo'
                                                                                           ELSE '... 275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0
2 12:08:30 SELECT * FROM activ_card
```

Una vez creada esta tabla la relaciono con Credit_cards

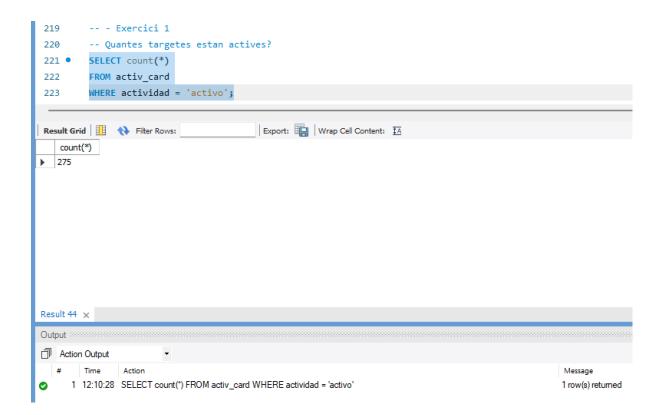
```
-- Relaciono esta tabla con el modelo
 216 • ALTER TABLE activ_card
         ADD CONSTRAINT fk_credit FOREIGN KEY(card_id) REFERENCES credit_cards(id);
 217
 218
Output :::
Action Output
        Time
                 Action
                                                                                                             Message

    14 11:21:54 SELECT * FROM transacciones.credit_cards

                                                                                                             275 row(s) returned
15 11:22:34 SELECT * FROM activ_card
                                                                                                             275 row(s) returned
16 11:26:42 ALTER TABLE activ_card ADD CONSTRAINT fk_credit FOREIGN KEY(card_id) REFERENCES credit_cards(id)
                                                                                                             275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0
```

- Exercici 1

Quantes targetes estan actives?



Nivel 3

Crea una tabla con la que podamos unir los datos del nuevo archivo products.csv con la base de datos creada, teniendo en cuenta que desde transaction tienes product_ids. Genera la siguiente consulta:

Se ha separado la columna products_ids de la tabla transactions, ya que en cada celda había los id de productos vendidos en una transacción separados por comas, entonces se han separado con Substring_index, respetando la transacción a la que pertenece, con esto se han creado una tabla llamada trans product.

```
CREATE TABLE trans_product AS
226 •
227
         SELECT id.
            SUBSTRING_INDEX(product_ids, ',', 1) AS producto
229
        FROM transactions
230
        WHERE product ids LIKE '%'
        UNION ALL
232
           SELECT id,
233
               SUBSTRING_INDEX(SUBSTRING_INDEX(product_ids, ',', -2), ',', -1) AS producto
           FROM transactions
234
235
            WHERE product_ids LIKE '%,%'
           UNION ALL
236
237
           SELECT id,
                 SUBSTRING_INDEX(SUBSTRING_INDEX(product_ids, ',', 2), ',', -1) AS producto
          FROM transactions
239
240
           WHERE product_ids LIKE '%,%,%'
           SELECT id.
242
243
                 SUBSTRING_INDEX(SUBSTRING_INDEX(product_ids, ',', -2), ',', 1) AS producto
           WHERE product ids LIKE '%,%,%,%';
245
246
        SELECT *
         FROM trans product;
248
Export: Wrap Cell Content: A Fetch rows:
                                       producto
  02C6201E-D90A-1859-B4EE-88D2986D3B02
                                      71
   0466A42E-47CF-8D24-FD01-C0B689713128
                                      47
   063FBA79-99EC-66FB-29F7-25726D1764A5
   0668296C-CDB9-A883-76BC-2E4C44F8C8AE
   06CD9AA5-9B42-D684-DDDD-A5E394FEBA99
   07A46D48-31A3-7E87-65B9-0DA902AD109F
   09DE92CE-6F27-2BB7-13B5-9385B2B3B8E2
trans_product 91 ×
Action Output
     1 20:31:17 CREATE TABLE trans_product AS SELECT id, SUBSTRING_INDEX(product_ids, ',', 1) AS producto FROM transacti...
                                                                                                     1457 row(s) affected Records: 1457 Duplicates: 0 Warnings: 0
2 20:31:17 SELECT * FROM trans_product
                                                                                                     1457 row(s) returned
```

Con esta nueva tabla creada, la relacionaremos con transactions y con products, para establecer una entidad – relación y que sirva de nexo o tabla puente para relacionarlas.

Así que modificamos el tipo de dato en la columna producto de la tabla trans_product a INT, para que se iguale en el tipo a la columna id de la tabla products.

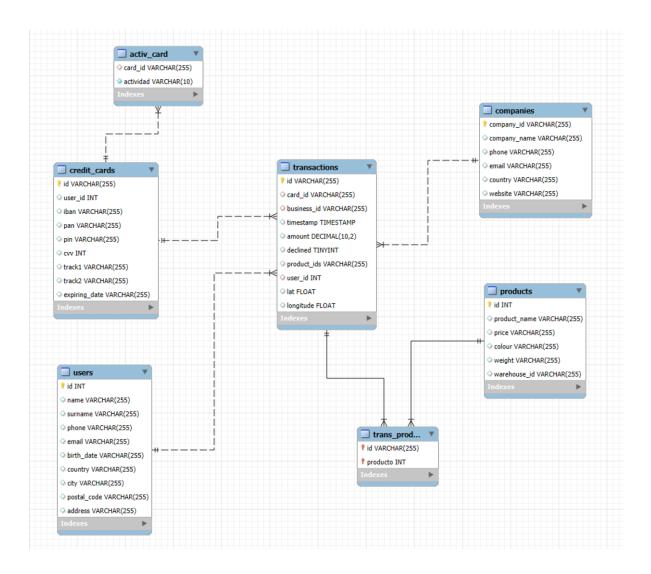
También creamos una primary key compuesta en la tabla trans_product con los dos campos que tiene esta tabla (id y producto)



Y finalmente, para relacionar Trans_product con trasactions y con products creamos las constraint de foreign keys.

```
ALTER TABLE trans_product MODIFY producto INT; -- modificar tipo de dato a INT
 253 •
 254
 255 •
          ALTER TABLE trans_product
           ADD CONSTRAINT fk_product FOREIGN KEY(producto) REFERENCES products(id);
 256
 257
 258 •
          ALTER TABLE trans_product
           ADD CONSTRAINT fk_id FOREIGN KEY (id) REFERENCES transactions(id);
 259
 261
Output
Action Output
    1 22:42:35 ALTER TABLE trans_product MODIFY producto INT
                                                                                                                  1457 row(s) affected Records: 1457 Duplicates: 0 Warnings: 0
2 22:42:35 ALTER TABLE trans_product ADD CONSTRAINT fk_product FOREIGN KEY(producto) REFERENCES products(id)
                                                                                                                  1457 row(s) affected Records: 1457 Duplicates: 0 Warnings: 0
      {\tt 3-22:42:36-ALTER\ TABLE\ trans\_product\ ADD\ CONSTRAINT\ fk\_id\ FOREIGN\ KEY\ (id)\ REFERENCES\ transactions (id)}
                                                                                                                  1457 row(s) affected Records: 1457 Duplicates: 0 Warnings: 0
```

Aquí muestro el diagrama final y sus relaciones.



Ejercicio 1

Necesitamos conocer el número de veces que se ha vendido cada producto.

Aquí mostramos cuantas veces se ha vendido cada producto,

Mostramos un listado de productos con la cantidad de ventas de cada producto que se ha vendido, excluimos las ventas hechas y que hayan sido rechazadas.

