SPRINT 4

Último repaso de SQL en el que trabajaremos la creación de bases de datos, las relaciones entre tablas, técnicas de modelado y granularidad de los datos.

Raquel Limpo Martínez

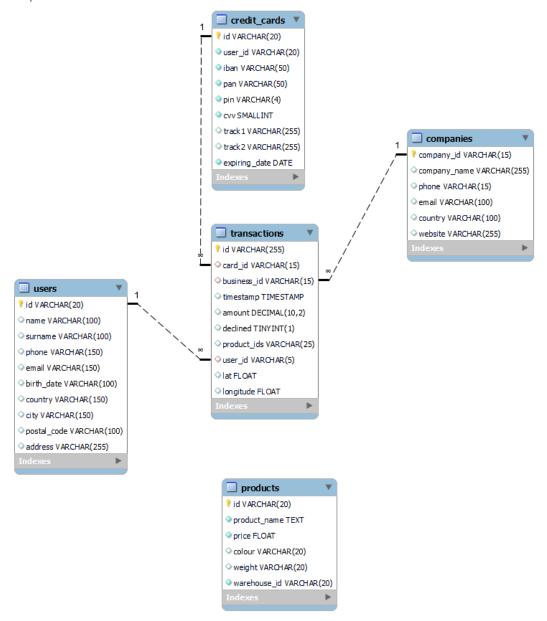
Tabla de contenido

Nivel 1	2
Esquema en estrella	2
Creación de la base de datos e importación de los datos de los CSV	5
Ejercicio 1: Realiza una subconsulta que muestre a todos los usuarios con más de 30 transacciones utilizando al menos 2 tablas.	11
Ejercicio 2 Muestra la media de amount por IBAN de las tarjetas de crédito a la compañ Donec Ltd., utiliza por lo menos 2 tablas.	
Nivel 2	13
Ejercicio 1: Crea una nueva tabla que refleje el estado de las tarjetas de crédito basado si las últimas tres transacciones fueron declinadas y genera la siguiente consulta: ¿Cuantas tarjetas están activas?	
Nivel 3	
Ejercicio 1: Necesitamos conocer el número de veces que se ha vendido cada producto	
Anexo: Los problemas de carga las tablas	20

Nivel 1

Descarga los archivos CSV, estúdialos y diseña una base de datos con un esquema en estrella que contengas al menos 4 tablas de las cuales puedas realizar las consultas de los ejercicios 1 y 2:

Esquema en estrella



En este esquema de estrella, la tabla factual o de hechos sería **Transaction**, no solo porque está conectada con todas las demás, sino que contiene los indicadores del negocio, cuanto se ha vendido, cuando a quien de parte de quien, cuánto dinero se ha ganado...

Está unida a las tablas de dimensiones: **Companies, Credit_cards y Users.** No me deja conectarla a products, porque la foreign key es muchos a muchos. Se soluciona en el nivel 3.

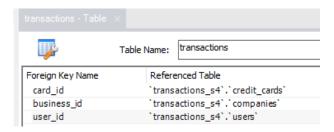


Tabla Transaction:

- id : El identificador de la transacción realizada. Es la clave primaria de esta tabla Como he comentado previamente, sería ideal que este campo fuera distintivo como *transaction_id*. Lo mismo aplica para el resto de tablas.
- card_id : El identificador de la tarjeta de crédito con la que se ha realizado la transacción, nos une a la tabla *Credit_card*
- business_id : El identificador de la compañía desde la que ha realizado la transacción, nos une a la tabla *Company*.
- timestamp : La hora a la que se ha realizado la transacción
- amount : La cantidad de dinero que se ha usado en la transacción
- declined : Campo booleano, nos indica si la transacción se ha realizado (0) o se ha cancelado (1)
- product_ids : los identificadores de los diferentes productos que se han intercambiado por dinero en la transacción. Cuando hay más de uno está separado por comas. Nos uniriamos a la tabla *Products* mediante una tabla intermedia que pondremos en el nivel 3..
- user_id : El identificador del usuario que ha realizado la transacción. Nos une por clave foránea a las tablas *Users*
- lat y longitude: Las coordenadas desde donde se ha hecho la transacción.

Tabla Company:

- company_id: El identificador para cada compañía. ES la clave primaria de esta tabla.
- company name. El nombre de la compañía.
- phone : El número de contacto de cada compañía.
- email: el email de contacto de cada compañia
- country: el país donde esta la sede central de cada compañía. Este sería el único campo que no tendría por qué ser único de esta tabla.
- website: La página oficial de la compañia.

Tabla Credit_cards: En esta tabla hay varios datos NOT NULL ya que he considerado que al ser importantes a la hora de hacer compra online era necesario sí o sí que estuvieran.

- id: Es el identificador de la tarjeta de crédito. Este campo es la clave primaria,
- iban: Es el código internacional de cuenta bancaria. Es un código alfanúmerico que identifica una cuenta bancaria en cualquier lugar del mundo.
- pan: es el acrónimo de personal account number y es el número que aparece en la tarjeta de crédito.
- pin: personal identification number es tu clave para entrar en tu cuenta.
- cvv: código valor de verificación. Es un número que está detrás de la tarjeta y que se suele pedir por seguridad en transacciones online.
- track1 y track2 : Son las pistas de banda magnética. Los lectores de losTPV siempre leen las pistas 1 o 2: la información mínima necesaria para realizar una transacción se encuentra en ambas pistas. La pista 1 tiene una mayor densidad de bits, es la única que puede contener caracteres alfanuméricos, y por tanto es la única que puede contener el nombre del portador de la tarjeta.
- Expiring_date La fecha de caducidad de la tarjeta. Al igual que el cvv se pide a menudo en transacciones online. Para cargarlo lo he hecho como VARCHARr, pero luego lo he editado a DATE

Tabla Product:

- id : El identificador del producto, es la primary key de la tabla.
- product_name : El nombre del producto.
- price El precio del producto, es un decimal. Lleva símbolo de dólar, con lo cual puede dar problema al cargar la tabla, pero se soluciona con un código extra como se ve más adelante.
- Colour: color en código hexadecimal del producto
- weight Peso del producto, sirve a la hora de enviar el producto y para tenerlo en cuenta al recibirlo.
- warehouse_id : Identificador del almacén, sirve para saber dónde está cada producto (qué almacén, qué zona del almacén, qué estante...).

Tablas Users de Canada, Reino Unido y Estados Unidos: todas tienen los mismos campos, así que las he unido todas en una sola tabla, para evitar repetición de código y simplificar las consultas.

- id: Identificador del usuario, clave primaria.
- Name: nombre del usuario
- Surname: apellido del usuario
- Phone: número de contacto del usuario
- Email: Correo electronico de contactodel usuario.
- birth_date : Fecha de nacimiento del usuario
- country: país de residencia del usuario
- city: Ciudad de residencia del usuario
- postal_code: Codigo postal del lugar de residencia del usuario
- address : dirección del lugar de residencia del usuario

Creación de la base de datos e importación de los datos de los CSV

Primero creamos la base de datos, para evitar problemas con bases de datos previas, le ponemos de nombre **transactions_s4**

Database y tabla companies

Creamos la tabla companies y la cargamos.

Para cargar los datos he buscado la ruta a través del código sql "SHOW VARIABLES LIKE 'secure_file_priv';", este te da la ruta exacta y te permite entrar en carpetas, aunque estén ocultas, pudiendo copiar ahí los archivos. Como estos dtos no tenian todos los campos de varchar entre comillas, he tenido que añadir "OPTIONALLY ENCLOSED BY "".

```
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/companies.csv'
INTO TABLE companies
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
```

Luego comprobamos que se han cargado los datos.

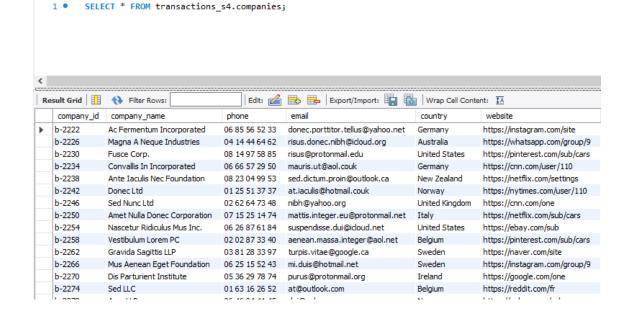


Tabla transactions

Repetimos el proceso con la tabla **transactions.** El reto con esta tabla fueron las foreign keys, ya que no deja hacer una relación de muchos a muchos con products_id.

En la parte de carga, además del problema con las comillas, se nos añadía el de que los campos no estaban separados por ',' sino por ';' .

```
# transactions

    ○ CREATE TABLE IF NOT EXISTS transactions (
      id VARCHAR(255) PRIMARY KEY,
      card_id VARCHAR(15) REFERENCES credit_card(id),
      business_id VARCHAR(15)REFERENCES companies(id),
      timestamp TIMESTAMP,
      amount DECIMAL(10,2),
      declined BOOLEAN,
      product_ids VARCHAR(25),
      user id VARCHAR(5) REFERENCES users(id),
      lat FLOAT,
      longitude FLOAT
      );
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/transactions.csv'
INTO TABLE transactions
FIELDS TERMINATED BY ';'
OPTIONALLY ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
 1 • SELECT * FROM transactions_s4.transactions;
```

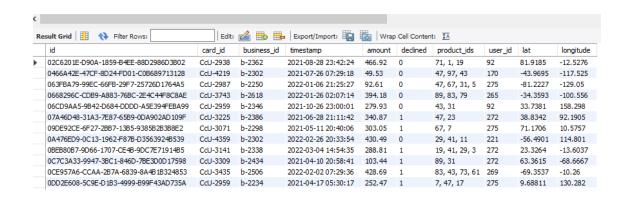


Tabla products

El precio estaba en dólares, así que no se cargaba correctamente, por lo que reemplacé el símbolo de dólar por un espacio usando el código:

```
(id, product_name, @price_raw, colour, weight, warehouse_id)
SET price = REPLACE(@price_raw, '$', '');.
```

```
#products
CREATE TABLE IF NOT EXISTS products (
     id VARCHAR(20) PRIMARY KEY,
     product_name TEXT NOT NULL ,
     price DECIMAL (10,2),
     colour VARCHAR(20),
     weight VARCHAR(20),
     warehouse_id VARCHAR(20) NOT NULL
· );
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/products.csv'
INTO TABLE products
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
TGNORE 1 ROWS
(id, product_name, @price_raw, colour, weight, warehouse_id)
SET price = REPLACE(@price raw, '$', '');
        SELECT * FROM transactions_s4.products;
```

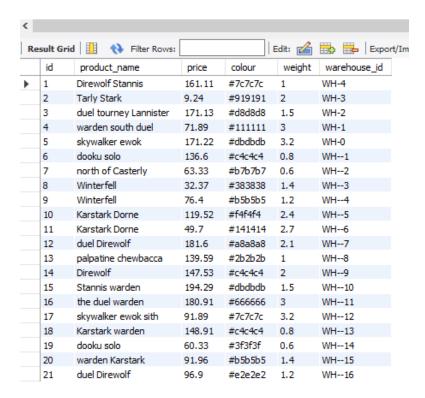


Tabla credit cards

Aquí el reto era cargar el campo expiring_date debido al estilo de las fechas, por lo que se ha cargado como VARCHAR y lego modificado a DATE. También estaba el error de las comillas que aparecen y desaparecen en algunos registros, pero como vimos anteriormente, se arregla con *OPTIONALLY ENCLOSED BY ""*

```
#credit cards
CREATE TABLE credit_cards (
         id VARCHAR(20) PRIMARY KEY,
         user_id VARCHAR(20) NOT NULL,
         iban VARCHAR(50) NOT NULL,
         pan VARCHAR(50) NOT NULL,
         pin VARCHAR(4) NOT NULL,
         CVV SMALLINT NOT NULL,
         track1 VARCHAR(255),
         track2 VARCHAR(255),
         expiring_date VARCHAR(20) NOT NULL
   );
 LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/credit_cards.csv'
 INTO TABLE credit cards
 FIELDS TERMINATED BY ','
 OPTIONALLY ENCLOSED BY '"'
 LINES TERMINATED BY '\n'
 IGNORE 1 ROWS:
 🚞 🔒 | 🗲 😿 👰 🕛 | 🗞 | 📀 🔞 | Don't Limit
                                                               - | 🏡 | 🍼 🔍 🗻 🖃
         SELECT * FROM transactions_s4.credit_cards;
                                           | Edit: 🚄 🖶 🖶 | Export/Import: 🏭 🐻 | Wrap Cell Content: 🖽
              user id iban
                                                                            CVV
                                                                                  track1
                                                                                                                         track2
                                                   5424465566813633
   CcU-2938
                      TR301950312213576817638661
                                                                                                                        %B7653863056044187=8007163336?3
   CdJ-2945 274
                     DO26854763748537475216568689
                                                  5142423821948828
                                                                     9080 887
                                                                                  %B4621311609958661^UftuyfsSeimxn^06106... %B4149568437843501=5107140330?1
                     BG45IVQL52710525608255
                                                                                  %B2183285104307501^CddyytcUxwfdq^5907... %B6778580257827162=69068597400?7
   CcU-2959 272
                     CR7242477244335841535
                                                  372461377349375
                                                                    3583
                                                                           667
                                                                                  %B7281111956795320^XocddijBckecd^09016... %B4246154489281853=2805223916?8
                                                  448566 886747 7265
                                                                                  %B4728932322756223^JhlgvsuFbmwgj^7202...
                     BG72LKTQ15627628377363
                                                                                                                        %B2318571115599881=8908215784?5
                                                                     4900
   CcU-2966
             271
                                                                           130
   CcU-2973 270
                     PT87806228135092429456346
                                                  544 58654 54343 384 8760
                                                                           887
                                                                                  %B4761405253275637^HjnnipoBlejrl^7108515... %B7816169831446746=13102772?9
                                                                                  %B7320483593870549^OokzqxrHpased^4901... %B2474313962214151=0412219131?5
   CcU-2980
                     DE39241881883086277136
                                                  402400 7145845969
                                                                     5075
                                                                           596
                                                  3763 747687 76666
   CcU-2987
                     GE89681434837748781813
                                                                                  %B4750646345146674^PjmlyrfGwwtrf^83051... %B5441935173418615=4103704536?7
                                                                     2298
   CcU-2994
             267
                     BH62714428368066765294
                                                  344283273252593
                                                                     7545
                                                                            595
                                                                                  %B1583759784015674^GmgoyhtUtogrn^2507... %B4141467473024349=65068009550?4
   CcU-3001
                     CY49087426654774581266832110
                                                  511722 924833 2244 9562
                                                                                  %B6227288756728648^AwxilfcFmgvdy^2808... %B3429355750963453=5305268305?3
                                                                                 %B7182449430529226^MkoutyhTfdvpo^1708... %B6235123731781366=940563816678
%B5776250106724742^QvvzkrcCwrovm^530... %B3561372148267521=1905473166?7
   CcU-3008
                     LU507216693616119230
                                                  4485744464433884
                                                                     1856
   CdU-3015
                     PS119398216295715968342456821 3784 662233 17389
   CcU-3022
             263
                     GT91695162850556977423121857
                                                  5164 1379 4842 3951
                                                                    5610
                                                                           342
                                                                                  %B8456622214332875^XmwxhenXwoxbt^050... %B7761154174374063=230423342974
   CcU-3029
                     AZ62317413982441418123739746
                                                  3429 279566 77631 9708
                                                                                  %B9882354238385082^HrmhqsnRprjpm^1902... %B4377026757181186=580245381976
                                                                                                                        %B7259434322726515=5305460495?3
   CcU-3036
                     AZ39336002925842865843941994
                                                  3768 451556 48766
                                                                     2232
                                                                            565
                                                                                  %B4814576454466431^IcstszwEpqgfk^92059...
   CcU-3043
             260
                     TN6488143310514852179535
                                                  455676 6437463635
                                                                           196
                                                                                  %B6524853618888394^VjjxpyyCbiopy^44067... %B5469323811685837=24038685?7
                                                                     5969
                                                                                                                                                         06,
                                                                                 %B7784177223537930^NhuwoziEexiwa^9105... %B2727867073512632=740369024475 %B4513246318178918^EbmpsehXroiiy^71025... %B1725272041127802=520276979872
   CcU-3050
                     FR5167744369175836831854477
                                                  4024007123722
                                                                     4834
   CcU-3057 258
                     LU931822574697545215
                                                  3484 621767 21237 6805 848
   CcU-3064 257
CcU-3071 256
                     PS146965545449253377627273133
                                                  3467 732741 26810
                                                                                  %B7675544533775797^XlxvwhuPwqeek^5103... %B2536323414661722=08123392228?
                                                  3464 789562 23352
                                                                    6625 661
                                                                                  %B8981277622416587^EhdwdwlMapqxc^280... %B8150826795832177=47081653206?1
                                                                                                                                                         12)
                     NO8923814763512
    CcU-3078
                      IS025127145884623279548733
                                                   4539 322 74 2377
                                                                                  %B8805395627676929^YnqcruwOohnno^690...
                                                  5266 3346 1135 1687 7241 413
   CcU-3085 254
                     BE63114723972437
                                                                                  %B8242982826686377^VnaovghMaifgu^7806... %B2848502349754061=46089897?9
                                                                                                                                                         05
                                                  3488 754223 46253
                     RO65LSOD1166122125447487
                                                                                  %B2804123689465462^DolminrDnylkw^61072...
   CcU-3099 252
                     PT26105275356823705537218
                                                  448 55418 98863 789 5612 564
                                                                                  %B5875431253326368^SdvavpeXpeeui^8403... %B7204024880651978=6509974687?4
                                                                                                                                                         01
   CcU-3106
                      AT684251637751136592
                                                  349547146395283
                                                                                  %B1491494879325988^FhfgvwhNxaoik^7001...
                                                                                                                        %B9821795210722817=03037207?7
                                                                     9733
                                                  341834822877471 9011 287
   CdU-3113 250
                     IE26LCGT47732173572752
                                                                                 %B9356731333359126^KwphyxlApsdpm^551... %B3438596027194536=5306586553?2
```

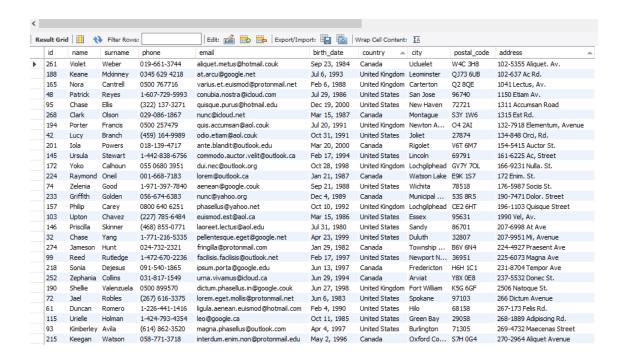
Tabla users:

He unido las tres tablas de users en una para evitar código repetido. En este caso el problema de carga era el salto de línea. Según el manual depende de qué documentos y cómo hubieran sido creados hacia falta poner el /r

http://download.nust.na/pub6/mysql/doc/refman/5.0/es/load-data.html#:~:text=Los%20valores%20FIELDS%20TERMINATED%20BY,BY%20'%5Cr%5Cn'%20.

"Nota: Si ha generado el fichero de texto en un sistema Windows, puede tener que usar LINES TERMINATED BY '\r\n' para leer correctamente el fichero, ya que los programas de Windows típicamente usan dos caracteres como terminadores de línea. Algunos programas como WordPad, pueden usar \r como terminador de línea al escribir ficheros. Para leer tales ficheros, use LINES TERMINATED BY '\r'."

```
■ ○ CREATE TABLE IF NOT EXISTS users (
        id VARCHAR(20) PRIMARY KEY,
        name VARCHAR(100),
        surname VARCHAR(100),
        phone VARCHAR(150),
        email VARCHAR(150),
        birth_date VARCHAR(100),
        country VARCHAR(150),
        city VARCHAR(150),
        postal code VARCHAR(100),
         address VARCHAR(255)
        );
 LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_ca.csv'
 INTO TABLE users
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY ""
LINES TERMINATED BY '\r\n'
 IGNORE 1 ROWS;
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_uk.csv'
INTO TABLE users
FIELDS TERMINATED BY ','
 OPTIONALLY ENCLOSED BY ""'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/users_usa.csv'
 INTO TABLE users
FIELDS TERMINATED BY ','
OPTIONALLY ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
 IGNORE 1 ROWS;
```

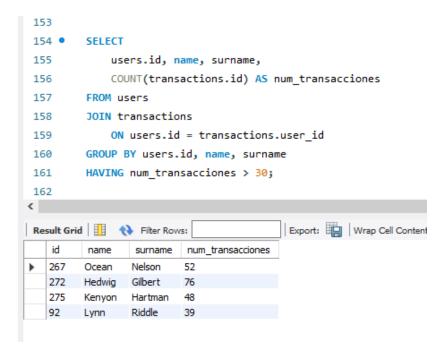


Ejercicio 1: Realiza una subconsulta que muestre a todos los usuarios con más de 30 transacciones utilizando al menos 2 tablas.

Para este ejercicio he pedido que me busque el nombre y apellidos de los usuarios y cuente la cantidad de transacciones realizadas, declinadas o no. He necesitado hacer una consulta con join de las tablas users y transactions y luego agruparlas por el dato único del id de los usuarios, usando el having para cribar aquellos que han hecho más de 30 transacciones.

Nos aparecen solo 4 usuarios con más de 30 transacciones.

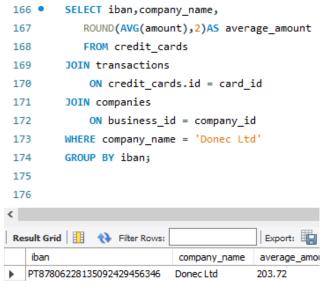
En un caso real debería preguntar al cliente si quería que fueran más de 30 o a partir de 30, aunque he comprobado que en este caso concreto no habría ninguna diferencia.



Ejercicio 2 Muestra la media de amount por IBAN de las tarjetas de crédito a la compañía Donec Ltd., utiliza por lo menos 2 tablas.

Buscamos la media del pago de la tarjeta de crédito y luego hacemos un doble join de credit cards con companies y transactions, después buscamos la compañía Donec y por último lo agrupamos por el Iban de la tarjeta de crédito.

La media resultante de Donec Ltd es de 203.72\$



Nivel 2

Ejercicio 1: Crea una nueva tabla que refleje el estado de las tarjetas de crédito basado en si las últimas tres transacciones fueron declinadas y genera la siguiente consulta: ¿Cuantas tarjetas están activas?

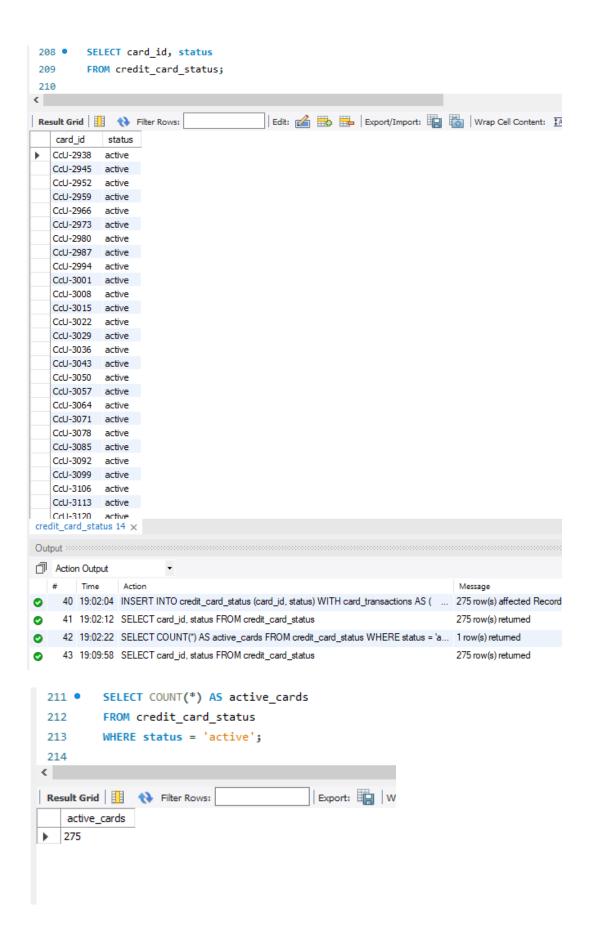
Creamos la tabla credit_card_status y le introducimos datos de las tablas credit_card y transactions, de la primera el campo de id y de la segunda obtenemos su estatus, si es activa o inactiva.

Para determinar el status usamos una subconsulta (WITH card_transactions) para identificar las últimas tres transacciones de cada tarjeta ordenadas por timestamp. Si la transacción ha sido declinada menos de 3 veces en las últimas 3 transacciones se convierte en activa. Después contamos cuantas tarjetas están en estado activo.

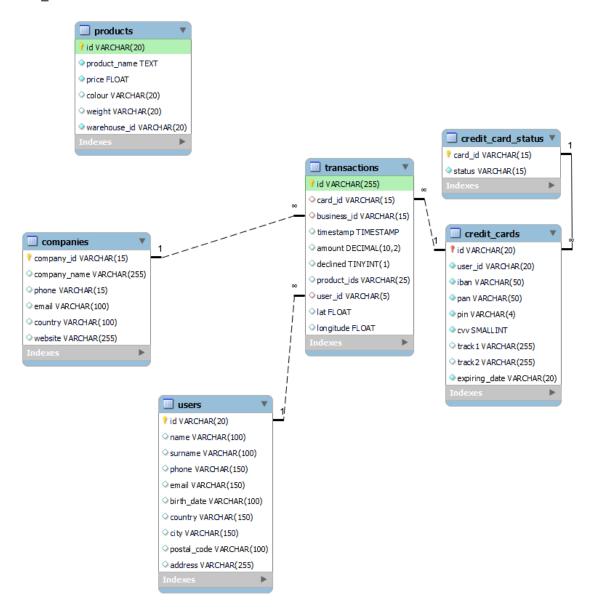
De 275 resultados no aparece ninguna tarjeta inactiva.

```
▶ ⊝ CREATE TABLE credit_card_status (
        card_id VARCHAR(15) PRIMARY KEY,
        status VARCHAR(15) NOT NULL,
        FOREIGN KEY (card_id) REFERENCES transactions(card_id)
    );
   CREATE INDEX idx_status ON credit_card_status (status);
  INSERT INTO credit_card_status (card_id, status)

⇒ WITH card_transactions AS (
     SELECT card_id,
            timestamp,
             declined.
             ROW_NUMBER() OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS row_transaction
      FROM transactions
  SELECT card_id,
            WHEN SUM(declined) <= 2 THEN 'active'
            ELSE 'inactive'
         END AS status
  FROM card transactions
  WHERE row_transaction <= 3
  GROUP BY card id;
```



El modelo relacional nos quedaría así, convirtiéndose el modelo estrella en uno de copo de nieve. La tabla credit_card status queda unida a credit_cards mediante la clave foránea de card_id.



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:

Creamos la tabla bridge_products con los campos transactions_id y products_id conectando así la tabla products con transactions.

Luego insertamos en la tabla bridge_products los id de transacciones y productos para todas las transacciones, eliminando los espacios en blanco del campo products_ids mediante la función FIND_IN_SET.

```
CREATE TABLE bridge_products (
    transactions_id VARCHAR(100),
    products_id VARCHAR(100),
    FOREIGN KEY (transactions_id) REFERENCES transactions(id),
    FOREIGN KEY (products_id) REFERENCES products(id)
);

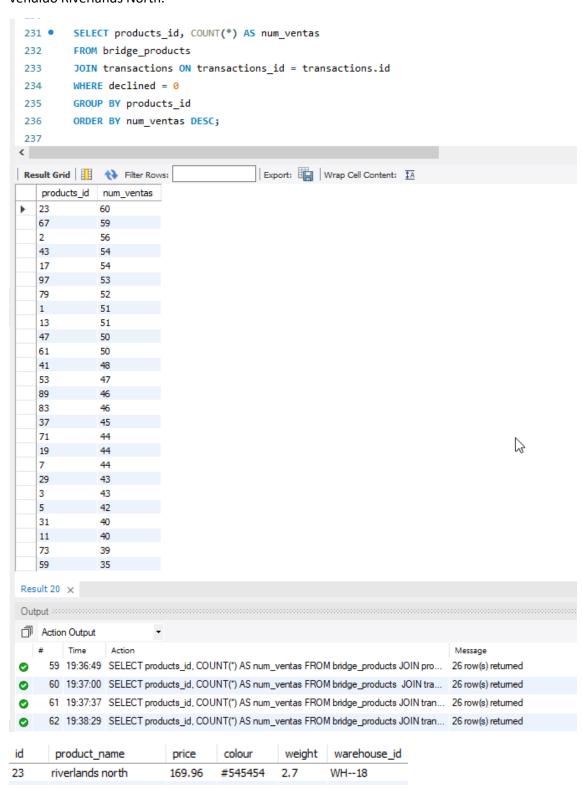
INSERT INTO bridge_products (transactions_id, products_id)
SELECT transactions.id AS transactions_id, products.id AS products_id
FROM transactions
JOIN products
ON FIND_IN_SET(products.id, REPLACE(transactions.product_ids, ' ', '')) > 0;
```

Comprobamos que los datos se han cargado correctamente en la tabla.

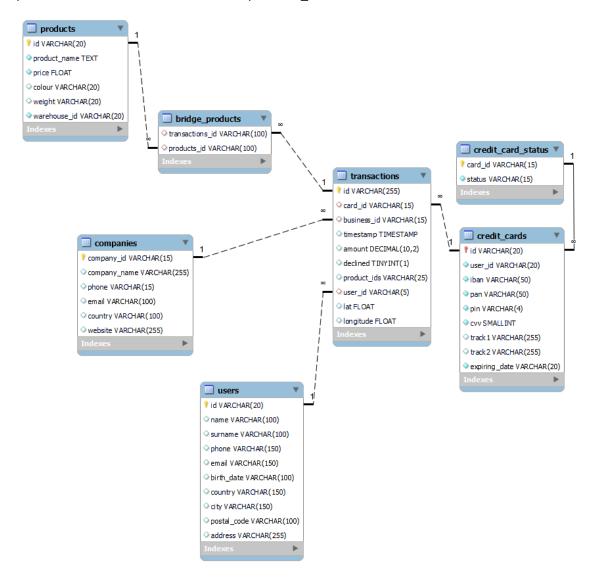


Ejercicio 1: Necesitamos conocer el número de veces que se ha vendido cada producto.

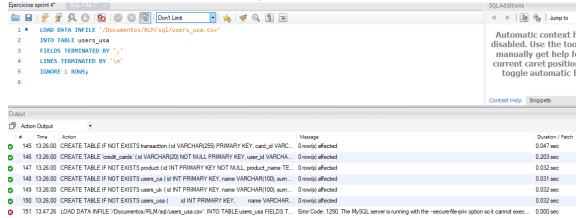
Nos da de resultado 26 columnas con el numero de ventas de cada producto-siendo el más vendido Riverlands North.



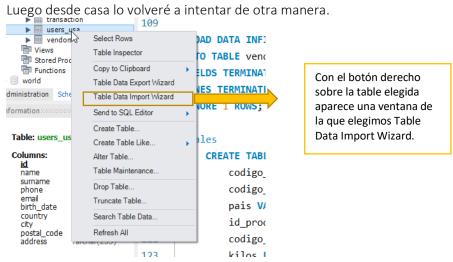
Por fin podemos unir products a transactions mediante la tabla bridge_products. Esta se une a products mediante la clave foránea de products_id.



Anexo: Los problemas de carga las tablas



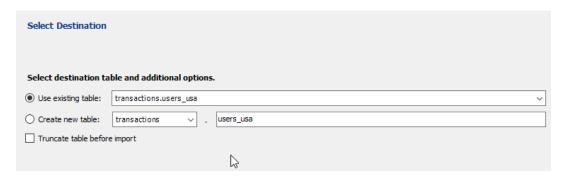
Originalmente iba a cargarlo así, però me aparecía el error "Error Code: 1290 - The MySQL server is running with the --secure-file-priv option so it cannot execute this statement." Y las maneras de corregirlo son difíciles en el ordenador del cibernàrium, ya que tengo que entrar en zonas seguras y no tengo los permisos, así que voy a ir avanzando, usando el import wizard.



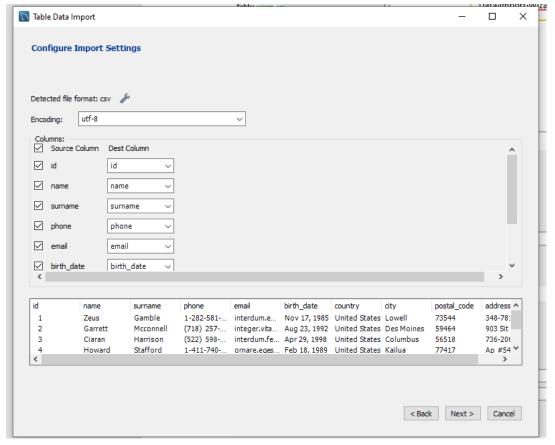
Nos abre el Wizard y le decimos de donde sacamos el archivo.

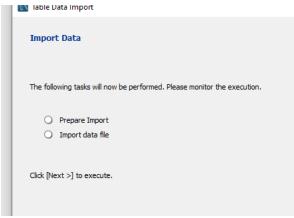


Le damos a Next y elegimos destino



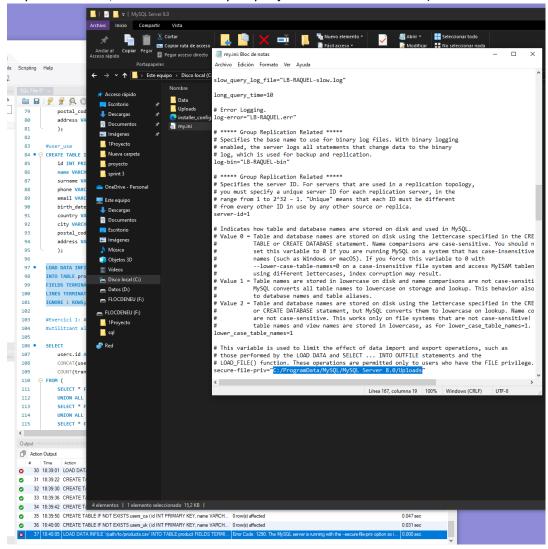
Le volvemos a dar a Next



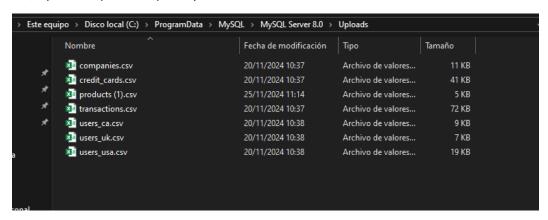


Como corregirlo desde casa:

Hay dos maneras, una abriendo el my.ini y dejando las comillas vacías (dónde esta resaltado).



La segunda, llevar los archivos csv a este directorio C:/ProgramData/MySQL/MySQL Server 8.0/Uploads , que es la que voy a hacer ahora.



Para cualquiera de ellas hay que volver a iniciar el mysql.