Sprint 4

Tasca S4.01. Creació de Base de Dades

Nivell 1

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:

Se crea la base de datos **transactionsnew** y cuatro tablas: *credit_cards, companies, users, transactions*.

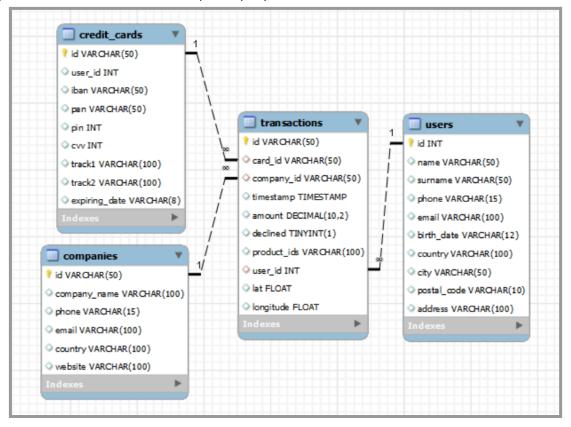
```
5 • ⊖ CREATE TABLE credit_cards (
        id VARCHAR(50) PRIMARY KEY,
        user_id INT,
        iban VARCHAR(50),
  9
        pan VARCHAR(50),
 10
       pin INT,
 11
        cvv INT,
        track1 VARCHAR(100),
        track2 VARCHAR(100),
 14
        expiring_date VARCHAR(8)
 15
 17 • ⊖ CREATE TABLE companies (
        id VARCHAR(50) PRIMARY KEY,
Output:
Action Output
      Time
     1 13:04:18 CREATE TABLE credit_cards (id VARCHAR(50) PRIMARY KEY, user_id INT, iban V... 0 row(s) affected
```

```
17 • ○ CREATE TABLE companies (
 18
          id VARCHAR(50) PRIMARY KEY,
          company_name VARCHAR(100),
 19
 20
          phone VARCHAR(15),
 21
          email VARCHAR(100),
          country VARCHAR(100),
 22
          website VARCHAR(100)
 23
          );
 24
Action Output
      1 13:04:18 CREATE TABLE credit_cards (id VARCHAR(50) PRIMARY KEY, user_id INT, iban V... 0 row(s) affected
Ø
      2 13:16:05 CREATE TABLE companies (id VARCHAR(50) PRIMARY KEY, company_name VAR... 0 row(s) affected
```

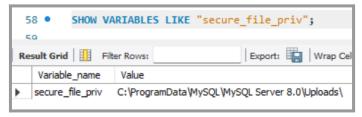
```
26 • ⊖ CREATE TABLE users (
 27
         id INT PRIMARY KEY,
         name VARCHAR(50),
 28
 29
         surname VARCHAR(50),
 30
         phone VARCHAR(15),
 31
         email VARCHAR(100),
         birth_date VARCHAR(12),
 32
 33
         country VARCHAR(100),
         city VARCHAR(50),
 34
         postal_code VARCHAR(10),
 35
         address VARCHAR (100)
 36
 37
Output :::::::
Action Output
      Time Action
                                                                                 Message
     1 13:17:58 CREATE TABLE users (id INT PRIMARY KEY, name VARCHAR(50), sumame VARC... 0 row(s) affected
```

```
39 • ○ CREATE TABLE transactions (
         id VARCHAR(50) PRIMARY KEY,
  40
  41
         card_id VARCHAR(50),
 42
         company_id VARCHAR(50),
 43
         timestamp TIMESTAMP,
         amount DECIMAL(10,2),
 44
         declined TINYINT(1),
 45
  46
         product_ids VARCHAR(100),
  47
         user_id INT,
  48
         lat FLOAT,
 49
         longitude FLOAT,
         FOREIGN KEY (card_id) REFERENCES credit_cards(id),
 50
         FOREIGN KEY (company_id) REFERENCES companies(id),
 51
         FOREIGN KEY (user_id) REFERENCES users(id)
 52
 53
 54
Output :::
Action Output
 # Time Action
                                                                                  Message
1 13:17:58 CREATE TABLE users (id INT PRIMARY KEY, name VARCHAR(50), sumame VARC... 0 row(s) affected
  2 13:23:42 CREATE TABLE transactions (id VARCHAR(50) PRIMARY KEY, card_id VARCHAR(... 0 row(s) affected, 1 warning(s): 1681 Integer display width is deprecated
```

Queda creado así el siguiente esquema de estrella con la tabla de hechos, *transactions*, como centro, y las tablas restantes como dimensiones. Las relaciones son de muchos a uno y relacionan las foreign keys de la tabla *transactions* con las primary keys de las tablas de dimensiones.



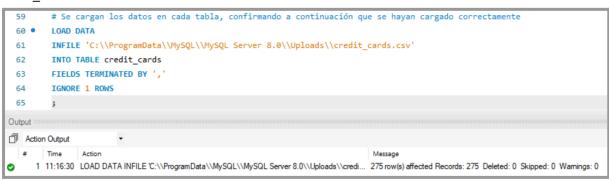
Para cargar los datos desde los archivos CSV con código SQL se colocan los archivos en la carpeta que MySQL habilita como fuente para cargar archivos (intenté modificar los permisos pero sin éxito). El siguiente código muestra la ubicación de la carpeta: SHOW VARIABLES LIKE "secure file priv"

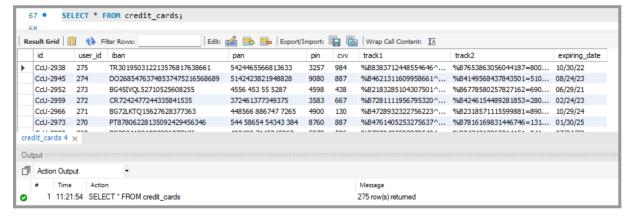


A la ruta del archivo se le agrega una backslash extra para que MySQL la lea correctamente. Se agregan las condiciones para que el CVS sea leído correctamente: el caracter que separa el campo es una coma o un punto y coma según el caso, en algunos archivos se debe especificar el final de línea con intro (\r) o nueva línea (\n). En todos los casos se ignora la primera fila, que son los nombres de los campos.

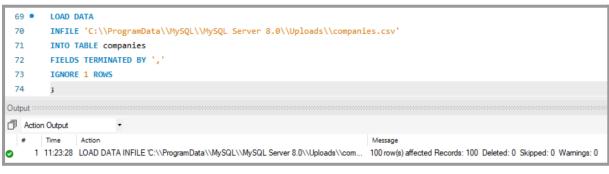
Se cargan los datos de los archivos CVS en las tablas respectivas y se confirma a continuación que se hayan cargado correctamente.

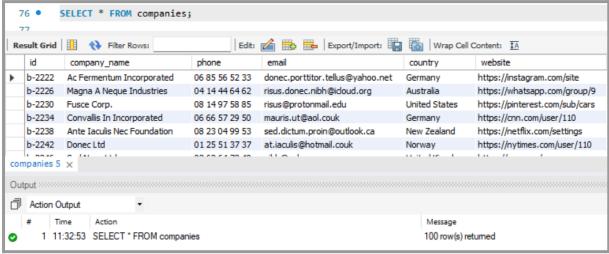
credit_cards



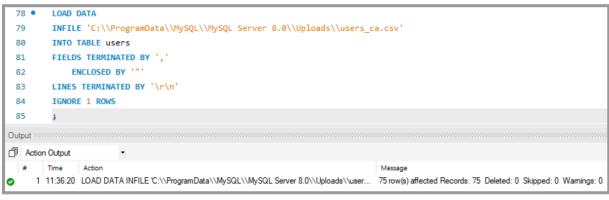


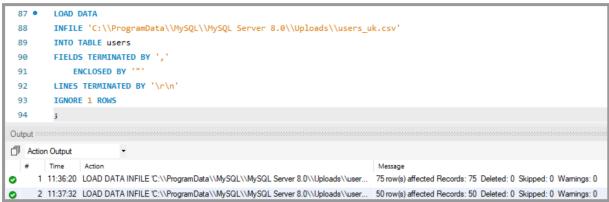
companies



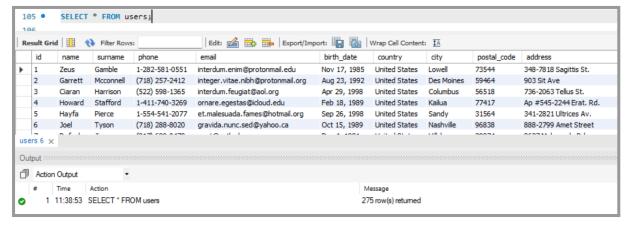


users



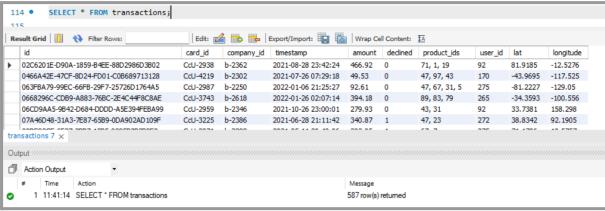






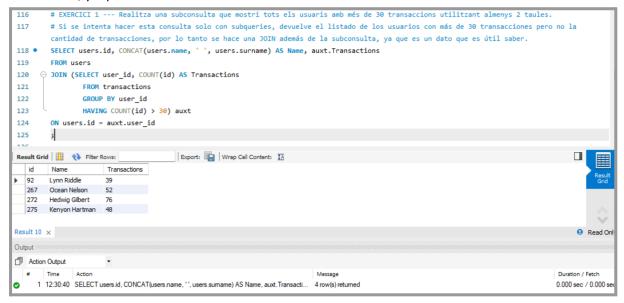
transactions





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

Si se intenta hacer esta consulta solo con subqueries, devuelve el listado de los usuarios con más de 30 transacciones pero no la cantidad de transacciones, por lo tanto se hace una JOIN además de la subconsulta, ya que es un dato úti.



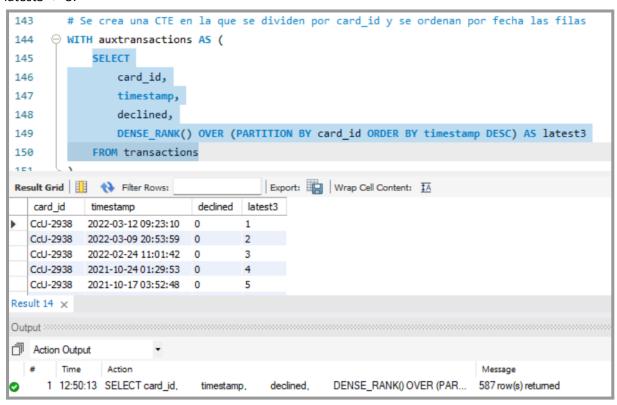
Mostra la mitjana d'amount per IBAN de les targetes de crèdit a la companyia Donec Ltd, utilitza almenys 2 taules.



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:

En el script de creación de la tabla se crea una CTE en la que se dividen por card_id y se ordenan por fecha las filas con la función DENSE_RANK() que en el campo *latest3* ordena las fechas de forma descendiente. Este campo más adelante se usa para filtrar las últimas tres transacciones indicando latest3 <= 3.



En la SELECT que usa la CTE auxtransactions se filtran las últimas tres transacciones en el WHERE y con CASE se asigna la etiqueta 'blocked' a las transacciones que tienen tres instancias en que declined sea True (es decir, en que la suma de los valores de declined es 3).

```
139
        # NIVELL 2
140
        # Crea una nova taula que reflecteixi l'estat de les targetes de crèdit basat en si les últimes tres transaccions van ser declinades
141 • CREATE TABLE card_status
142
143
       # Se crea una CTE en la que se dividen por card_id y se ordenan por fecha las filas
145
           SELECT
146
               card_id,
147
               timestamp,
148
               declined,
               DENSE_RANK() OVER (PARTITION BY card_id ORDER BY timestamp DESC) AS latest3
149
            FROM transactions
150
151
152
     # En la SELECT que usa la CTE auxtransactions en su FROM, se filtran las últimas tres transacciones en el WHERE y con CASE se asigna la
        etiqueta 'active' a las transacciones que no tienen ninguna instancia declined (que la suma de los valores de decline es 0 en las tres
        transacciones)
153
      SELECT
154
            card_id,
155 ⊝
Output ::
Action Output
                                                                                                                                       Duration / Fetch
1 13:06:54 CREATE TABLE card_status AS # Se crea una CTE en la que se dividen por card_id ... 275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0
                                                                                                                                       0.047 sec
```

```
153
       SELECT
154
           card_id,
155
           CASE
               WHEN SUM(declined) = 3 THEN 'blocked'
156
157
               ELSE 'active'
158
           END AS status
159
       FROM auxtransactions
160
       WHERE latest3 <= 3
        GROUP BY card_id
161
        5
162
```

Luego se agrega la foreign key para vincular la tabla a transactions.

```
# Se agrega la foreign key para vincular la tabla a transactions

165 • ALTER TABLE card_status

166 ADD FOREIGN KEY (card_id) REFERENCES credit_cards(id)

167 

Output

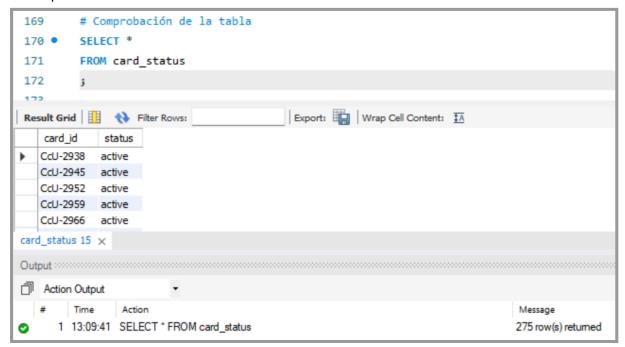
Action Output

# Time Action

Action Action Action Action Message

1 13:09:11 ALTER TABLE card_status ADD FOREIGN KEY (card_id) REFERENCES credit_card... 275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0
```

Se comprueba la tabla.



Quantes targetes estan actives?

Todas las tarjetas están activas.

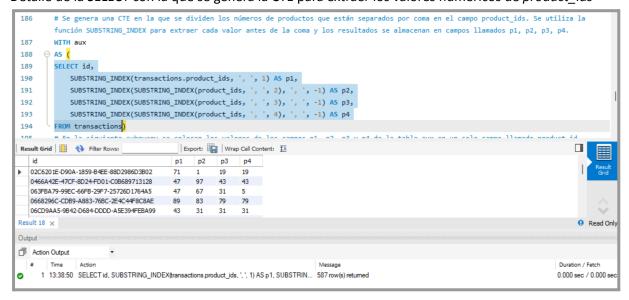


Nivell 3

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:

- Se crea una tabla puente llamada bridge_products con los campos *transaction_id* (vinculado con el campo *id* de la tabla transactions) y *product_id* (vinculado con el campo *id* de la tabla products).
- Se genera una CTE en la que se dividen los números de productos que están separados por coma en el campo *product_ids*. Se utiliza la función SUBSTRING_INDEX para extraer cada valor antes de la coma y los resultados se almacenan en campos llamados *p1*, *p2*, *p3*, *p4*.
- En la siguiente subquery se colocan los valores de los campos *p1*, *p2*, *p3* y *p4* de la tabla aux en un solo campo llamado *product_id* utilizando la función UNION, y no Union All, para no incluir los valores repetidos (que se generaron en el paso anterior, pues SUBSTRING_INDEX devuelve el último valor si no encuentra uno nuevo). Entiendo que generar valores repetidos no es la forma más eficiente de código, pero considero que al ser una CTE que se genera una sola vez para crear una tabla, no es un uso constante de recursos.
- Se ordena por *id* para mayor claridad al visualizar la tabla.
- Se crea la tabla products y se agregan las foreign keys para vincular las tablas transactions y products a través de la tabla puente.
- Se introducen los datos en tabla products.

Detalle de la SELECT con la que se genera la CTE para extraer los valores numéricos de product_ids



Query completa de la creación de la tabla bridge_products

```
tabla products)
184 •
         CREATE TABLE bridge_products (transaction_id VARCHAR(50), product_id INT)
185
         # Se genera una CTE en la que se dividen los números de productos que están separados por
186
         coma en el campo product ids. Se utiliza la función SUBSTRING INDEX para extraer cada valor
         antes de la coma y los resultados se almacenan en campos llamados p1, p2, p3, p4.
         WITH aux
187
188
      SELECT id,
189
190
             SUBSTRING INDEX(transactions.product ids, ', ', 1) AS p1,
             SUBSTRING_INDEX(SUBSTRING_INDEX(product_ids, ', ', 2), ', ', -1) AS p2,
191
             SUBSTRING_INDEX(SUBSTRING_INDEX(product_ids, ', ', 3), ', ', -1) AS p3,
192
             SUBSTRING INDEX(SUBSTRING INDEX(product ids, ', ', 4), ', ', -1) AS p4
193
       FROM transactions)
194
195
         # En la siguiente subquery se colocan los valores de los campos p1, p2, p3 y p4 de la tabla
         aux en un solo campo llamado product_id utilizando la función UNION, y no UNION ALL, para
         no incluir los valores repetidos (que se generaron en el paso anterior, pues
         SUBSTRING INDEX devuelve el último valor si no encuentra uno nuevo)
196
         SELECT id AS transaction_id, product_id
197

⊖ FROM (SELECT id, p1 AS product_id)

198
             FROM aux
199
            LINTON
200
            SELECT id, p2
            FROM aux
201
202
            UNION
203
            SELECT id, p3
           FROM aux
204
205
            UNION
            SELECT id, p4
206
207
             FROM aux
208
       ) AS aux2
209
         # Se ordena por id para mayor claridad al visualizar la tabla
210
211
212
213
         # Se crea la tabla products
Output seess
Action Output
                                                                                              Duration / Fetch
     1 08:39:51 CREATE TABLE bridge_products (transaction_id VA... 1457 row(s) affected Records: 1457 Duplicates: 0 ...
                                                                                              0.094 sec
```

Creación de la tabla products

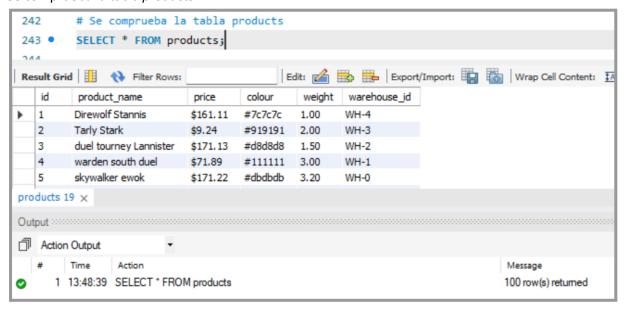
```
214
         # Se crea la tabla products
215 ● ⊖ CREATE TABLE products (
216
         id TNT PRIMARY KEY.
         product_name VARCHAR(100),
217
        price VARCHAR(20),
218
        colour VARCHAR(20),
219
        weight DECIMAL(8,2),
220
221
         warehouse_id VARCHAR(15)
       );
222
Output ::
Action Output
               Action
                                                                                 Message
    1 13:44:15 CREATE TABLE products (id INT PRIMARY KEY, product_name VARCHAR(100), pri... 0 row(s) affected
```

Se agregan las FK para vincular las tablas transactions y products a través de la tabla puente

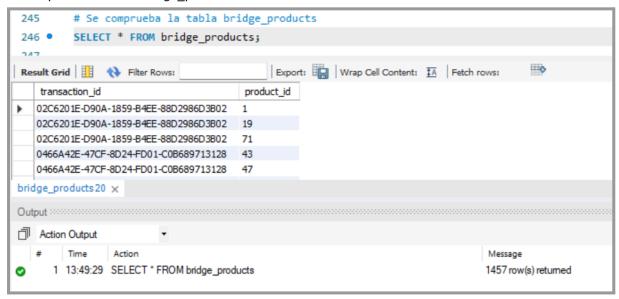
Se introducen los datos en la tabla products

```
234
         # Se introducen los datos en la tabla products
235 • LOAD DATA
236
        INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\products.csv'
237
        INTO TABLE products
238
        FIELDS TERMINATED BY ','
       IGNORE 1 ROWS
239
240
Output ::::
Action Output
 # Time
              Action
                                                                               Message
1 13:47:38 LOAD DATA INFILE 'C:\\ProgramData\\MySQL\\MySQL Server 8.0\\Uploads\\prod... 100 row(s) affected Records: 100 Deleted: 0 Skipped: 0 Warnings: 0
```

Se comprueba la tabla products



Se comprueba la tabla bridge_products



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

