

# SPRINT 4 rev2

DATA ANALYST

RAFAEL FONS BIBILONI

04-Junio-2025

## NIVEL I

Siguiendo las recomendaciones, primero creo las estructuras de tablas, cargo los datos y luego establezco las relaciones entre tablas.

```
11 ##### Nivel 1 #####
12 /*
13 Descarga los archivos CSV, estudiales y diseña una base de datos con un esquema de estrella que contenga,
14 al menos 4 tablas de las que puedas realizar las siguientes consultas:
15 */
16 • CREATE DATABASE sprint4;
17 • USE sprint4;
18
19 -- primero creo la tabla de dimensiones "users":
20 • CREATE TABLE users (
21     id INT PRIMARY KEY,
22     name VARCHAR(50),
23     surname VARCHAR(50),
24     phone VARCHAR(50),
25     email VARCHAR(50),
26     birth_date VARCHAR(20),
27     country VARCHAR(50),
28     city VARCHAR(100),
29     postal_code VARCHAR(50),
30     address VARCHAR(200)
31 );
32
```

100% 1:15

Action	Output
Time	Action
6	14:30:08 CREATE DATABASE sprint4
7	14:30:11 USE sprint4
8	14:30:16 CREATE TABLE users ( id INT PRIMARY KEY, name VARCHAR(50), surname VARCHAR(50), phone VARCHAR(50), email VARCHAR(50), birth_date... 0 row(s) affected

Cargo los datos en “users”. Para poder importar los ficheros “users\_XX.csv” tuve que ajustar los parámetros de la sintaxis LOAD INFILE:

- Primero agregar la línea “ENCLOSED BY ‘’” porque el campo birth\_date tiene “,” en su texto.
- Luego, modifiqué el tipo de salto de línea “LINES TERMINATED BY ‘\r\n’ para evitar problemas entre Mac y Windows:

```
33 -- Inserto los datos/registros en la tabla "users":
34 • LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/users_uk.csv'
35 INTO TABLE users
36 FIELDS TERMINATED BY ','
37 ENCLOSED BY ''
38 LINES TERMINATED BY '\r\n'
39 IGNORE 1 ROWS
40 (id, name, surname, phone, email, birth_date, country, city, postal_code, address);
41
42 • LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/users_ca.csv'
43 INTO TABLE users
44 FIELDS TERMINATED BY ','
45 ENCLOSED BY ''
46 LINES TERMINATED BY '\r\n'
47 IGNORE 1 ROWS
48 (id, name, surname, phone, email, birth_date, country, city, postal_code, address);
49
50 • LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/users_usa.csv'
51 INTO TABLE users
52 FIELDS TERMINATED BY ','
53 ENCLOSED BY ''
54 LINES TERMINATED BY '\r\n'
55 IGNORE 1 ROWS
56 (id, name, surname, phone, email, birth_date, country, city, postal_code, address);
57
```

00% 7:29

Action	Output
Time	Action
9	14:32:30 LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/da... 50 row(s) affected Records: 50 Deleted: 0 Skipped: 0 Warnings: 0
10	14:32:38 LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/da... 75 row(s) affected Records: 75 Deleted: 0 Skipped: 0 Warnings: 0
11	14:32:41 LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/da... 150 row(s) affected Records: 150 Deleted: 0 Skipped: 0 Warnings: 0

# SPRINT 4 rev2

DATA ANALYST

RAFAEL FONS BIBILONI

04-Junio-2025

Creo la tabla “credit\_cards” y cargo los datos del fichero csv.

```
58 -- creo la tabla de dimensiones "credit_cards":
59 • CREATE TABLE credit_cards (
60     id VARCHAR(8) PRIMARY KEY,
61     user_id INT,
62     iban VARCHAR(50),
63     pan VARCHAR(50),
64     pin INT,
65     cvv INT,
66     track1 VARCHAR(100),
67     track2 VARCHAR(100),
68     expiring_date DATE
69 );
70
71 -- cargo los datos en la tabla "credit_cards":
72 • LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/credit_cards.csv'
73 INTO TABLE credit_cards
74 FIELDS TERMINATED BY ','
75 LINES TERMINATED BY '\n'
76 IGNORE 1 ROWS
77 (id, user_id, iban, pan, pin, cvv, track1, track2, @expiring_date)          -- el @ me almacena temporalmente el valor y luego lo carga
78 SET expiring_date = STR_TO_DATE(@expiring_date, '%m/%d/%y');                -- indico el formato en el que esta la fecha en el fichero csv
79
```

Action Output

Time	Action	Response
12 14:33:41	CREATE TABLE credit_cards ( id VARCHAR(8) PRIMARY KEY, user_id INT, iban VARCHAR(50), pan VARCHAR(50), pin INT, cvv INT, tr...	0 row(s) affected
13 14:33:45	LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/da...	275 row(s) affected Records: 275 Deleted: 0 Skipped: 0 Warnings...

Creo la tabla “companies” y cargo los datos del fichero csv.

```
80 -- creo la tablas "companies" y cargo los datos:
81 • CREATE TABLE companies (
82     company_id VARCHAR(6) PRIMARY KEY,
83     company_name VARCHAR(50),
84     phone VARCHAR(20),
85     email VARCHAR(50),
86     country VARCHAR(20),
87     website VARCHAR(50)
88 );
89
90 -- cargo los datos en la tabla "companies":
91 • LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/companies.csv'
92 INTO TABLE companies
93 FIELDS TERMINATED BY ','
94 LINES TERMINATED BY '\n'
95 IGNORE 1 ROWS
96 (company_id, company_name, phone, email, country, website);
97
```

Action Output

Time	Action	Response
14 14:36:25	CREATE TABLE companies ( company_id VARCHAR(6) PRIMARY KEY, company_name VARCHAR(50), phone VARCHAR(20), email VAR...	0 row(s) affected
15 14:36:28	LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/da...	100 row(s) affected Records: 100 Deleted: 0 Skipped: 0 Warnings...

# SPRINT 4 rev2

DATA ANALYST

RAFAEL FONS BIBILONI

04-Junio-2025

Creo la tabla “products” y cargo los datos del fichero csv.

```
97 -- creo la tablas "products" y cargo los datos:
98 • CREATE TABLE products (
99     id INT PRIMARY KEY,
100    product_name VARCHAR(100),
101   price DECIMAL(10,2),
102   colour VARCHAR(10),
103   weight DECIMAL(5,1),
104   warehouse_id VARCHAR(10)
105 );
106
107 • LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/products.csv'
108 INTO TABLE products
109 FIELDS TERMINATED BY ','
110 LINES TERMINATED BY '\n'
111 IGNORE 1 ROWS
112 (id,product_name,@price,colour,weight,warehouse_id)
113 SET price = CAST(REPLACE(@price, '$', '') AS DECIMAL(10,2));
114
100% 7:92
Action Output
Time Action Response
19 14:41:28 CREATE TABLE products ( id INT PRIMARY KEY, product_name VARCHAR(100), price DECIMAL(10,2), colour VARCHAR(10), weight D... 0 row(s) affected
20 14:41:31 LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/products.csv' 100 row(s) affected Records: 100 Deleted: 0 Skipped: 0 Warnings: 0
```

Creo la tabla “transactions” y cargo los datos del fichero csv.

```
115 -- finalmente creo la tabla de hechos principal "transactions";
116 • CREATE TABLE transactions (
117     id VARCHAR(150) PRIMARY KEY,
118     card_id VARCHAR(8),
119     business_id VARCHAR(6),
120     timestamp DATETIME,          -- uso DATETIME porque el formato de los datos cumple con YYYY-MM-DD HH:MM:SS
121     amount DECIMAL(10,2),
122     declined INT,
123     product_ids VARCHAR(30),
124     user_id INT,
125     lat DECIMAL(20,18),
126     longitude FLOAT
127 );
128
129 • LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/transactions.csv'
130 INTO TABLE transactions
131 FIELDS TERMINATED BY ';'
132 ENCLOSED BY ''
133 LINES TERMINATED BY '\r\n'
134 IGNORE 1 ROWS
135 (id, card_id, business_id, timestamp, amount, declined, product_ids, user_id, lat, longitude);
136
00% 24:110
Action Output
Time Action Response
21 14:43:44 CREATE TABLE transactions ( id VARCHAR(150) PRIMARY KEY, card_id VARCHAR(8), business_id VARCHAR(6), timestamp DA... 0 row(s) affected
22 14:43:47 LOAD DATA INFILE '/Users/rafafons/Library/Mobile Documents/com~apple~CloudDocs/2_Rafa/01-Capacitaciones/IT Academy/sprint 4/datos/transactions.csv' 587 row(s) affected Records: 587 Deleted: 0 Skipped: 0 Warnings: 0
```

Creo ahora las relaciones de la tabla principal “transactions” con el resto de las tablas (excepto “products”):

```
137 -- creo las relaciones con las demas tablas
138 • ALTER TABLE transactions
139 ADD FOREIGN KEY (card_id) REFERENCES credit_cards(id),
140 ADD FOREIGN KEY (user_id) REFERENCES users(id),
141 ADD FOREIGN KEY (business_id) REFERENCES companies(company_id);
142
100% 13:129
Action Output
Time Action Response
23 14:45:21 ALTER TABLE transactions ADD FOREIGN KEY (card_id) REFERENCES credit_cards(id), ADD FOREIGN KEY (user_id) REFERENCES us... 587 row(s) affected Records: 587 Duplicates: 0 Warnings: 0
```

# SPRINT 4 rev2

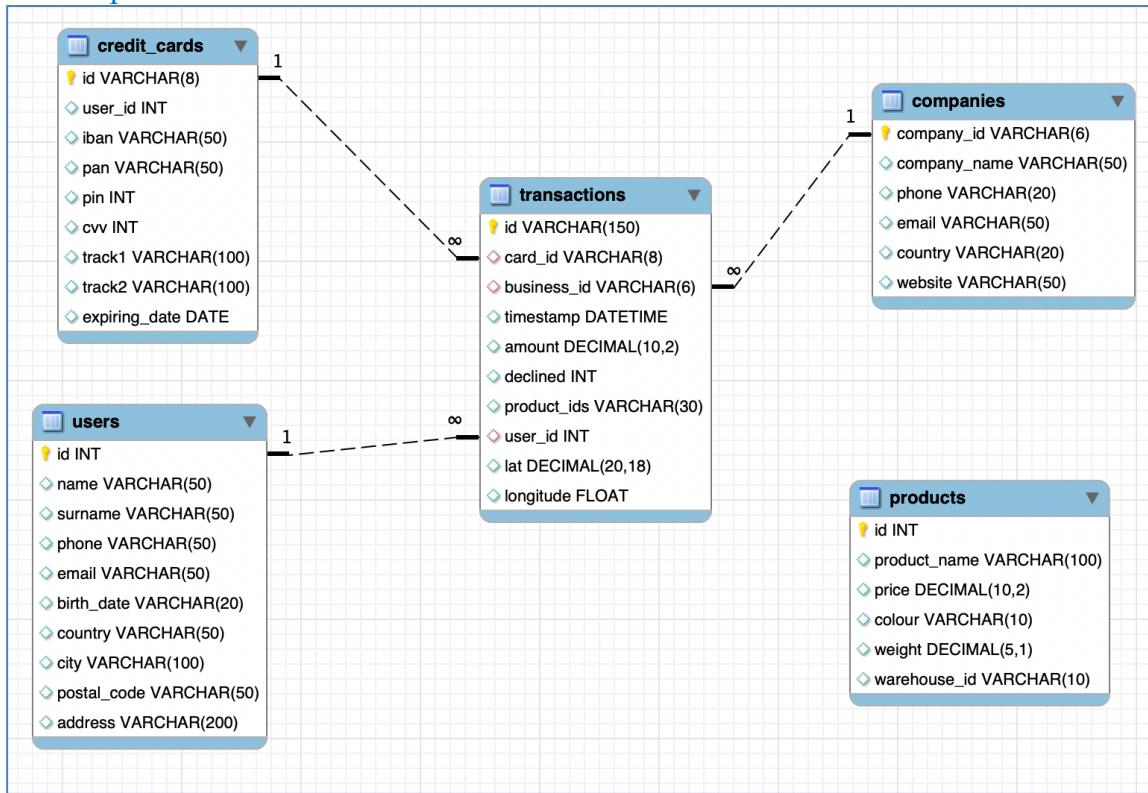
DATA ANALYST

RAFAEL FONS BIBILONI

04-Junio-2025

El modelo me queda de la siguiente manera. Donde hay que revisar varios aspectos:

- La tabla “products” no está relacionada con ninguna tabla.
- La línea discontinua refiere a que son relaciones blandas o no restrictas, es decir puedo dejar esos campos en NULL.



## NIVEL I - EJERCICIO 1:

Hago ahora la consulta solicitada haciendo una subconsulta en el JOIN:

```
148 -- hago la consulta que muestre a todos los usuarios con más de 30 transacciones:  
149 • SELECT u.id, u.name, u.surname, t.cant_transaction  
150   FROM users u  
151   JOIN (  
152     SELECT user_id, COUNT(*) AS cant_transaction  
153     FROM transactions  
154     WHERE declined = 0  
155     GROUP BY user_id  
156     HAVING COUNT(*) > 30  
157   ) t ON u.id = t.user_id;  
158
```

Result Grid Filter Rows: Search Export:

id	name	surname	cant_transaction
92	Lynn	Riddle	39
267	Ocean	Nelson	39
272	Hedwig	Gilbert	38

Result 7

Action Output

Time	Action	Response
36 15:08:26	SELECT u.id, u.name, u.surname, t.cant_transaction FROM users u JOIN ( SELECT user_id, COUNT(*) AS cant_transaction FROM... 3 row(s) returned	

# SPRINT 4 rev2

DATA ANALYST

RAFAEL FONS BIBILONI

04-Junio-2025

## NIVEL I - EJERCICIO 2:

Hago la consulta solicitada, mostrando la media de “amount” por IBAN de las tarjetas de crédito en la compañía Donec Ltd.

```
158  /*
159   NIVEL I - Ejercicio 2
160   Muestra la media de amount por IBAN de las tarjetas de crédito en la compañía Donec Ltd., utiliza por lo menos 2 tablas
161 */
162 -- Hago la consulta solicitada, mostrando la media de amount por IBAN de las tarjetas de crédito en la compañía Donec Ltd:
163 • SELECT c.company_name, cd.iban, ROUND(AVG(t.amount),2) AS prom_amount
164   FROM transactions t
165   JOIN credit_cards cd ON t.card_id = cd.id
166   JOIN companies c ON t.business_id = c.company_id
167   WHERE c.company_name = "Donec Ltd"
168   GROUP BY cd.iban;
169
00% 12:152
```

Result Grid Filter Rows: Search Export:

company_name	iban	prom_amount
Donec Ltd	PT87806228135092429456346	203.72

Result 8

Action Output

Time	Action	Response
37 15:10:48	SELECT c.company_name, cd.iban, ROUND(AVG(t.amount),2) AS prom_amount FROM transactions t JOIN credit_cards cd ON t.card_id = cd.id WHERE c.company_name = "Donec Ltd"	1 row(s) returned

## NIVEL II - EJERCICIO 1:

Creo la tabla intermedia entre “transactions” y “products”. Luego inserto los datos: activa = 1 e inactiva = 0.

```
171  /*
172   NIVEL II - Ejercicio 1
173   Crea una nueva tabla que refleje el estado de las tarjetas de crédito basado en si las últimas tres transacciones fueron declinadas
174   y genera la siguiente consulta: ¿Cuántas tarjetas están activas?
175 */
176 -- Creo la nueva tabla "active_cards"
177 • CREATE TABLE cards_status(
178   card_id VARCHAR(8),
179   card_status BOOL,
180   FOREIGN KEY (card_id) REFERENCES credit_cards(id)
181 );
182
183 -- inserto los datos en la tabla "cards_status"
184 • INSERT INTO cards_status (card_id, card_status)
185   SELECT
186     cd.id AS card_id,
187     IF(
188       (SELECT COUNT(*)
189        FROM (
190          SELECT t2.declined
191            FROM transactions t2
192            WHERE t2.card_id = cd.id
193            ORDER BY t2.timestamp DESC
194            LIMIT 3
195          ) temp
196          WHERE temp.declined = 1
197        ) = 3, -- si las ultimas 3 transacciones tienen un declined =1, entonces tarjeta INACTIVE
198        0, -- caso si verdadero (las 3 ultimas son declinadas)
199        1 -- caso si falso (las 3 ultimas no son declinadas)
200      ) AS card_status
201   FROM credit_cards cd;
202
203 -- visualizo los datos:
00% 25:193
```

Action Output

Time	Action	Response
42 23:43:27	CREATE TABLE cards_status( -- no tendría mas sentido crear una vista? ya que sería algo dinámico en lugar de estático card_id... 0 row(s) affected	
43 23:43:40	INSERT INTO cards_status (card_id, card_status) SELECT cd.id AS card_id, IF( -- uso un IF para que verifique la condición e... 275 row(s) affected Records: 275 Duplicates: 0 Warnings: 0	

# SPRINT 4 rev2

DATA ANALYST

RAFAEL FONS BIBILONI

04-Junio-2025

Visualizo los datos de la tabla recién creada:

```
203 -- visualizo los datos:  
204 • SELECT * FROM cards_status;  
205
```

100% 18:198

Result Grid		Filter Rows:	Search	Export:
cards_status 10				
card_id	card_status			
CcU-2943	1			
CcU-2952	1			
CcU-2959	1			
CcU-2966	1			
CcU-2973	1			
...	...			
cards_status 10				
Action Output				
	Time	Action		Response
44	23:47:01	SELECT * FROM cards_status LIMIT 0, 50000		275 row(s) returned

Hago la consulta para ver la cantidad de tarjetas activas:

```
206 -- Finalmente, respondo a la consulta, de cuantas tarjetas "activas" (card_status=1) hay:  
207 • SELECT count(*)  
208   FROM cards_status  
209   WHERE card_status = 1;  
210
```

100% 28:204

Result Grid		Filter Rows:	Search	Export:
count(*)				
275				
Result 11				
Action Output				
	Time	Action		Response
45	23:48:01	SELECT count(*) FROM cards_status WHERE card_status = 1 LIMIT 0, 50000		1 row(s) returned

## NIVEL III - EJERCICIO 1:

Creo la tabla “transaction\_products” y cargo los datos de los ids usando la función FIND\_IN\_SET:

```
211 ##### Nivel 3 #####  
212 • /*  
213   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  
214   transaction tienes product_ids.  
215 */  
216   -- primero creare una tabla donde recogeré los campos: transaction.id y transaction_products_ids:  
217 • CREATE TABLE transaction_products (  
218     transaction_id VARCHAR(150),  
219     product_id INT  
220 );  
221  
222 • INSERT INTO transaction_products (transaction_id, product_id)  
223   SELECT t.id AS transaction_id, p.id AS product_id  
224   FROM transactions t  
225   JOIN products p  
226   ON FIND_IN_SET(CAST(p.id AS CHAR), REPLACE(t.product_ids, ',', ''))>0;  
227
```

100% 22:203

Action Output		Time	Action	Response
50	23:53:19	CREATE TABLE transaction_products ( transaction_i...	0 row(s) affected	
51	23:53:24	INSERT INTO transaction_products (transaction_id, pr...	1457 row(s) affected Records: 1457 Duplicates: 0 Warnings: 0	

# SPRINT 4 rev2

DATA ANALYST

RAFAEL FONS BIBILONI

04-Junio-2025

Agrego las relaciones con las FK:

```
229 -- Agrego en la tabla "transaction_products" las claves foráneas a FK transactions.id y products.id:  
230 • ALTER TABLE transaction_products  
231 ADD FOREIGN KEY (transaction_id) REFERENCES transactions(id),  
232 ADD FOREIGN KEY (product_id) REFERENCES products(id);  
233
```

100% 1:210

Action Output

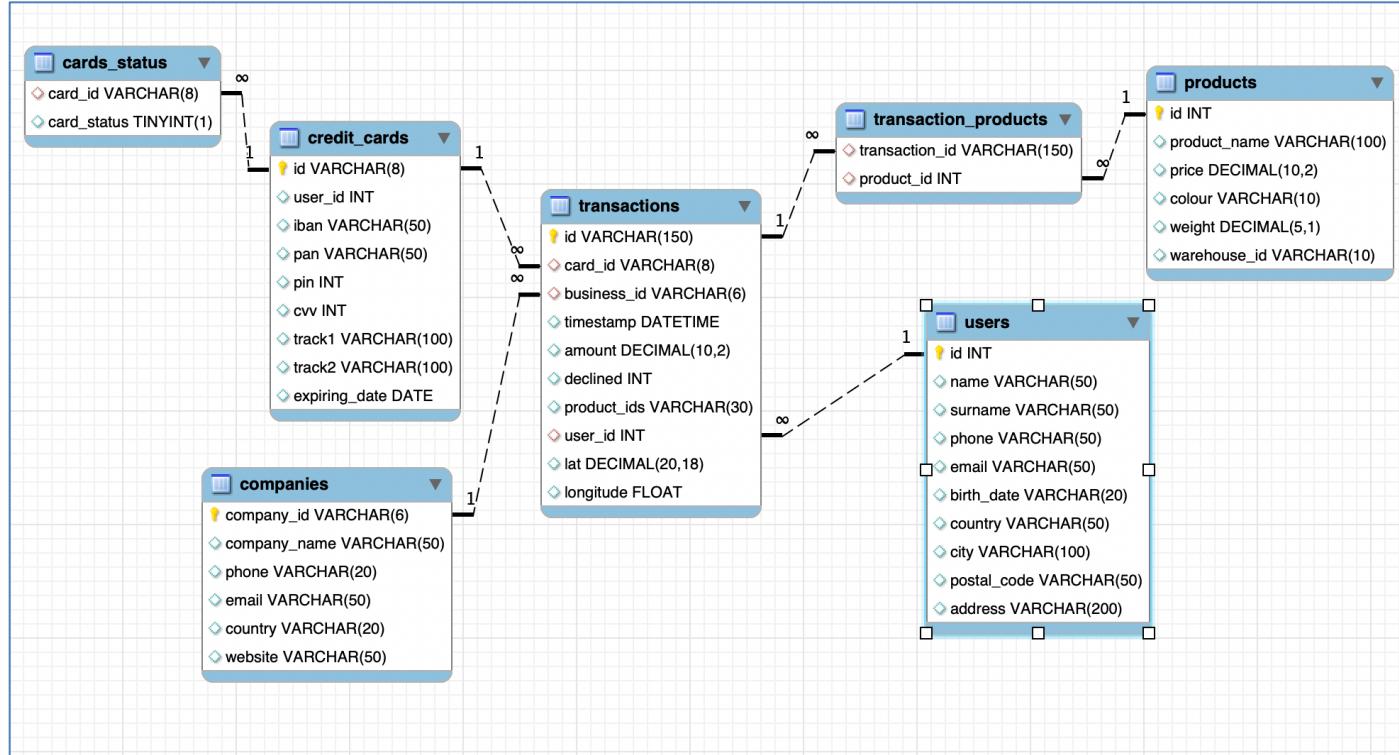
Time

Action

Response

52 23:53:52 ALTER TABLE transaction\_products ADD FOREIGN KEY... 1457 row(s) affected Records: 1457 Duplicates: 0 Warnings: 0

El modelo queda ahora con la siguiente estructura de tipo “snowflake”:



Hacemos finalmente la consulta solicitada:

```
235 /*  
236 Nivel III - Ejercicio 1  
237 Genera la siguiente consulta: Necesitamos conocer el número de veces que se ha vendido cada producto.  
238 */  
239  
240 • SELECT tp.product_id, p.product_name, count(*) AS cant_vendida  
241 FROM transaction_products tp  
242 JOIN transactions t ON tp.transaction_id = t.id  
243 JOIN products p ON tp.product_id = p.id  
244 WHERE t.declined = 0  
245 GROUP BY tp.product_id  
246 ORDER BY cant_vendida DESC;  
247
```

100% 30:231

Result Grid Filter Rows: Search Export:

product_id	product_name	cant_vendida
23	riverlands north	60
67	Winterfell	59
2	Tarly Stark	56
43	duel	54
17	shewalker awok with	54

Result 12

Action Output

Time

Action

Response

53 23:56:00 SELECT tp.product\_id, p.product\_name, count(\*) AS c... 26 row(s) returned