Tasca S4.01. Manipulació de taules
Descripció
Partint d'alguns arxius CSV dissenyaràs i crearàs la teva 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:

Exercici 1
Realitza una subconsulta que mostri tots els usuaris amb més de 80 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.

Després de fer una ullada a la informació dels documents csv proporcionats, he creat la base de dades 's4_transactions' i les taules que contindran la informació de cadascun dels csv. Les taules les he anomenat de la mateixa manera que els documents csv per no generar confusió. Per tal de no tenir problemes a l'hora d'importar les dades totes les columnes estan en format VARCHAR. El codi es veu a les captures a continuació:

```
-- DROP DATABASE s4_transactions;
CREATE DATABASE IF NOT EXISTS s4_transactions; -- AÑADIDO 'IF NOT EXISTS' P2P

SHOW DATABASES; -- AÑADIDO P2P

USE s4_transactions;

/*
CREATE TABLE IF NOT EXISTS american_users ( -- AÑADIDO 'IF NOT EXISTS' P2P
    id VARCHAR (200),
    name VARCHAR (200),
    surname VARCHAR (200),
    phone VARCHAR (200),
    email VARCHAR (200),
```

Result Grid | Filter Rows: Search | Export:

Database
information_schema
mysql
performance_schema
s4_transactions
sys
transactions

Result 19 | Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
24	13:08:52	CREATE DATABASE IF NOT EXISTS s4_transactions	1 row(s) affected	0.0020 sec
25	13:08:52	SHOW DATABASES	6 row(s) returned	0.0021 sec / 0.00001...
26	13:08:52	USE s4_transactions	0 row(s) affected	0.00063 sec
27	13:08:52	CREATE TABLE IF NOT EXISTS american_users (-- AÑADIDO 'IF NOT EXISTS' P2P id VARCHAR (200), name VAR...	0 row(s) affected	0.014 sec

```
CREATE TABLE IF NOT EXISTS american_users ( -- AÑADIDO 'IF NOT EXISTS' P2P
    id VARCHAR (200),
    name VARCHAR (200),
    surname VARCHAR (200),
    phone VARCHAR (200),
    email VARCHAR (200),
    birth_date VARCHAR (200),
    country VARCHAR (200),
    city VARCHAR (200),
    postal_code VARCHAR (200),
    address VARCHAR (200)
    );

CREATE TABLE IF NOT EXISTS european_users ( -- AÑADIDO 'IF NOT EXISTS' P2P
    id VARCHAR (200),
    name VARCHAR (200),
    surname VARCHAR (200),
    phone VARCHAR (200),
    email VARCHAR (200),
    birth_date VARCHAR (200),
    country VARCHAR (200),
    city VARCHAR (200),
    postal_code VARCHAR (200),
    address VARCHAR (200)
    );
```

```
CREATE TABLE IF NOT EXISTS credit_cards ( -- AÑADIDO 'IF NOT EXISTS' P2P
    id VARCHAR (200),
    user_id VARCHAR (200),
    iban VARCHAR (200),
    pan VARCHAR (200),
    pin VARCHAR (200),
    cvv VARCHAR (200),
    track1 VARCHAR (200),
    track2 VARCHAR (200),
    expiring_date VARCHAR (200)
    );

CREATE TABLE IF NOT EXISTS transactions ( -- AÑADIDO 'IF NOT EXISTS' P2P
    id VARCHAR (200),
    card_id VARCHAR (200),
    business_id VARCHAR (200),
    timestamp VARCHAR (200),
    amount VARCHAR (200),
    declined VARCHAR (200),
    product_ids VARCHAR (200),
    user_id VARCHAR (200),
    lat VARCHAR (200),
    longitude VARCHAR (200)
    );
```

```sql
CREATE TABLE IF NOT EXISTS companies ( -- AÑADIDO 'IF NOT EXISTS' P2P
company_id VARCHAR (200),
company_name VARCHAR (200),
phone VARCHAR (200),
email VARCHAR (200),
country VARCHAR (200),
website VARCHAR (200)
);

CREATE TABLE IF NOT EXISTS products ( -- AÑADIDO 'IF NOT EXISTS' P2P
id VARCHAR (200),
product_name VARCHAR (200),
price VARCHAR (200),
colour VARCHAR (200),
weight VARCHAR (200),
warehouse_id VARCHAR (200)
);
```

Per tal de poder importar els csv vaig canviar la configuració d'activar local_infile així com canviar una connexió del Workbench per tal que reconegués el path del document (codi a la captura posat com a comentari).

```sql
/*
-- Activate local_infile so we can add csv from local
SHOW GLOBAL VARIABLES LIKE 'local_infile';
SET GLOBAL local_infile = ON;
SET PERSIST local_infile = ON;
SHOW GLOBAL VARIABLES LIKE 'local_infile';

-- Insert data from csv files
SHOW VARIABLES LIKE 'secure_file_priv';
It gaves me an error when I tried to load data. I followed the following advice and htne it worked:
2068 no és del servidor: és el client (MySQL Workbench / driver) que rebutja enviar el fitxer local.
Solució a Workbench (la que acostuma a arreglar-ho)
Database → Manage Connections…
Selecciones la connexió → Edit
Pestanya Advanced
A Others afegeix:
OPT_LOCAL_INFILE=1
*/
```

A continuació vaig importar cadascun dels csv amb LOAD DATA LOCAL INFILE a cadascuna de les taules creades, definint paràmetres extra com el separador entre columnes i com acabava cada línia. El path complet s'ha tapat en la captura.

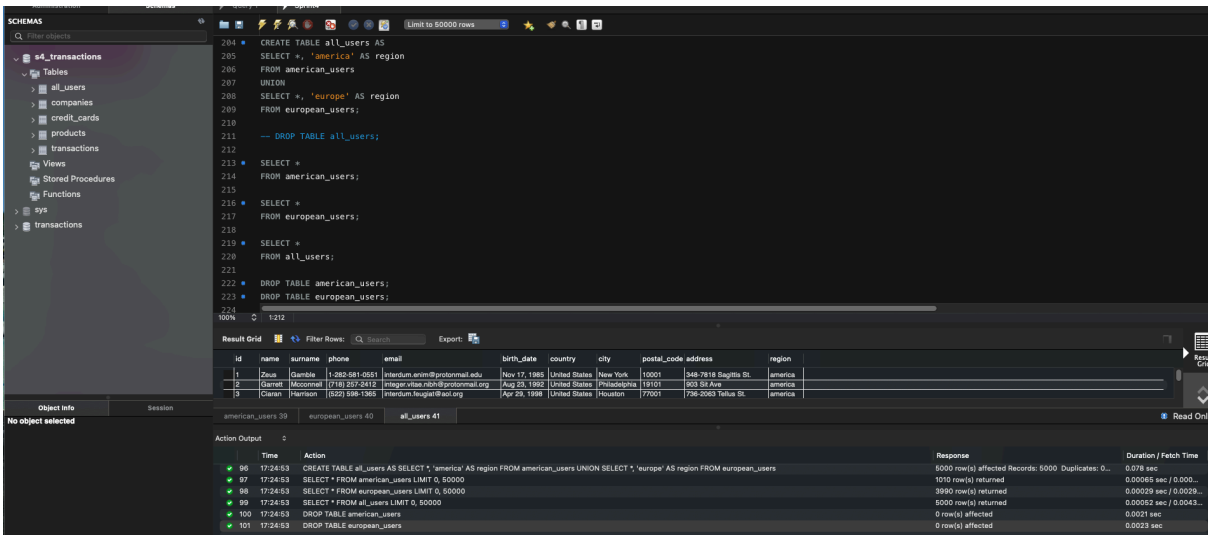
```
LOAD DATA LOCAL INFILE PATH data/american_users.csv'
INTO TABLE american_users
FIELDS TERMINATED BY ','
ENCLOSED BY '"' -- birth_date is enclosed by "", it gave a warning since it divided its info by 2 (column1: "Nov 17 column2: 1985"), I did select * from the table. I used TRUNCATE TALE to del
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

-- TRUNCATE TABLE american_users;

SELECT * from american_users
LIMIT 2;


LOAD DATA LOCAL INFILE PATH /data/companies.csv'
INTO TABLE companies
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

-- TRUNCATE TABLE companies

SELECT * from companies
```

	Time	Action	Response	Duration / Fetch Time
66	17:06:38	LOAD DATA LOCAL INFILE PATH data/american_users.csv' INTO TABLE american_users FIELDS TERMINAT...	1010 row(s) affected Records: 1010 Deleted: 0 Skipp...	0.022 sec
67	17:06:38	SELECT * from american_users LIMIT 2	2 row(s) returned	0.00036 sec / 0.000...
68	17:06:38	LOAD DATA LOCAL INFILE PATH data/companies.csv' INTO TABLE companies FIELDS TERMINATED BY ',' L...	100 row(s) affected Records: 100 Deleted: 0 Skippe...	0.0024 sec
69	17:06:38	SELECT * from companies LIMIT 2	2 row(s) returned	0.00021 sec / 0.0000...
70	17:06:38	LOAD DATA LOCAL INFILE PATH data/credit_cards.csv' INTO TABLE credit_cards FIELDS TERMINATED BY...	5000 row(s) affected Records: 5000 Deleted: 0 Skip...	0.044 sec
71	17:06:38	SELECT * from credit_cards LIMIT 2	2 row(s) returned	0.00048 sec / 0.000...
72	17:06:38	LOAD DATA LOCAL INFILE PATH data/european_users.csv' INTO TABLE european_users FIELDS TERMINA...	3990 row(s) affected Records: 3990 Deleted: 0 Skip...	0.030 sec
73	17:06:38	SELECT * from european_users LIMIT 2	2 row(s) returned	0.00037 sec / 0.0000...
74	17:06:38	LOAD DATA LOCAL INFILE PATH data/products.csv' INTO TABLE products FIELDS TERMINATED BY ';' LINE...	100 row(s) affected Records: 100 Deleted: 0 Skippe...	0.0014 sec
75	17:06:38	SELECT * from products LIMIT 2	2 row(s) returned	0.00017 sec / 0.0000...
76	17:06:38	LOAD DATA LOCAL INFILE PATH data/transactions.csv' INTO TABLE transactions FIELDS TERMINATED BY...	100000 row(s) affected Records: 100000 Deleted: 0...	0.334 sec
77	17:06:38	SELECT * from transactions LIMIT 2	2 row(s) returned	0.00037 sec / 0.0000...

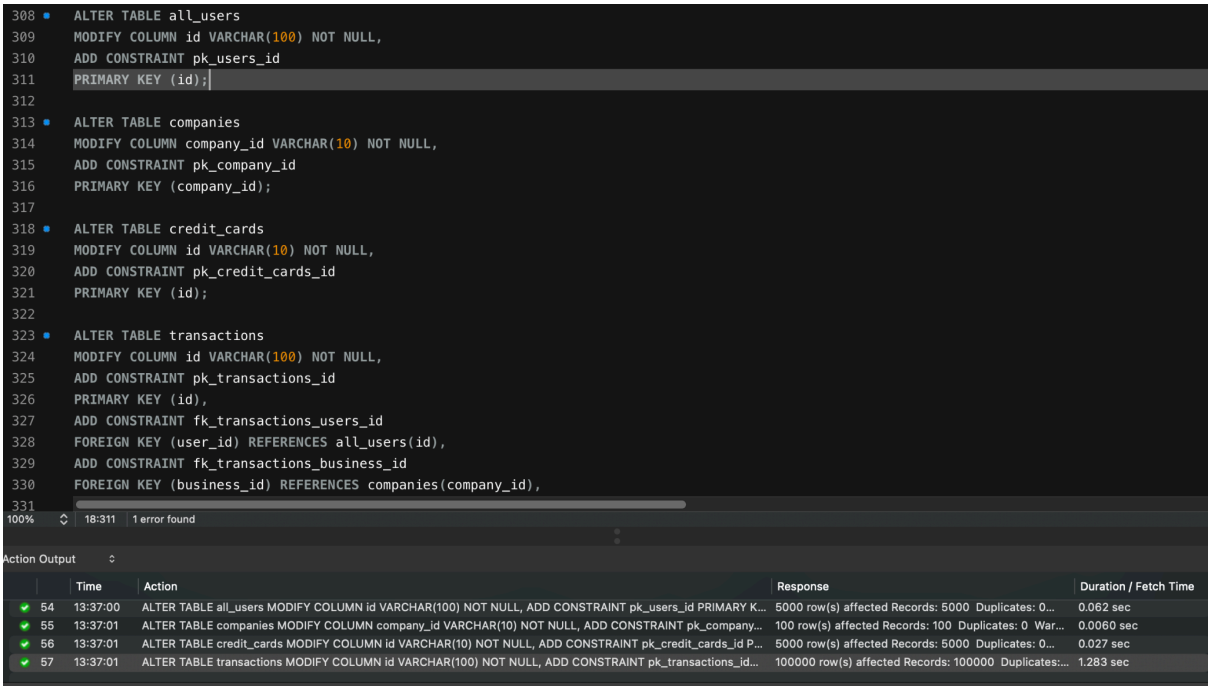
```
LOAD DATA LOCAL INFILE PATH data/credit_cards.csv'
INTO TABLE credit_cards
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

-- TRUNCATE TABLE credit_cards

SELECT * from credit_cards
LIMIT 2;


LOAD DATA LOCAL INFILE PATH data/european_users.csv'
INTO TABLE european_users
FIELDS TERMINATED BY ','
ENCLOSED BY '"' -- birth_date is enclosed by "", it gave a warning since it divided its info by 2 (column1: "Nov 17 column2: 1985"),
LINES TERMINATED BY '\n'
IGNORE 1 LINES;
```

```
LOAD DATA LOCAL INFILE PATH /data/products.csv'
INTO TABLE products
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

-- TRUNCATE TABLE products

SELECT * from products
LIMIT 2;


LOAD DATA LOCAL INFILE PATH 'data/transactions.csv'
INTO TABLE transactions
FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

-- TRUNCATE TABLE transactions
```

A continuació, he ajuntat en una sola taula ‘all_users’ tots els usuaris provinents de les taules ‘american_users’ i ‘european_users’, creant una nova columna amb la informació d’on prové cada entrada. Prèviament havia comprovat visualment que els números no coincidien i un cop fet, que hi havia el mateix nombre d’entrades a l’hora de fer el UNION (no UNION ALL per veure si es perdien, sumant les dues taules dóna el mateix nombre d’entrades que la taula final). Finalment he eliminat les taules inicials dels usuaris separats.



Després he afegit les primary i foreign keys a les taules que s'utilitzaran per a les consultes que es demanen així com per a fer el diagrama. També he visualitzat alguna de les taules per a saber exactament què hi havia a dintre de cada taula.




```
ALTER TABLE transactions
MODIFY COLUMN id VARCHAR(100) NOT NULL,
ADD CONSTRAINT pk_transactions_id
PRIMARY KEY (id),
ADD CONSTRAINT fk_transactions_users_id
FOREIGN KEY (user_id) REFERENCES all_users(id),
ADD CONSTRAINT fk_transactions_business_id
FOREIGN KEY (business_id) REFERENCES companies(company_id),
ADD CONSTRAINT fk_transactions_credit_cards_id
FOREIGN KEY (card_id) REFERENCES credit_cards(id);

SELECT *
FROM transactions;

SELECT *
FROM companies;

SELECT *
FROM credit_cards;
```

Finalment he realitzat les dues consultes que es demanaven (l'enunciat està a l'inici d'aquest document així com al codi) i el diagrama de les relacions entre les diverses taules, que està en format estrella. La taula productes no l'he relacionat perquè s'ha de crear una taula outrigger i no era necessària per aquestes consultes.

```
/*Exercici 1
Realitza una subconsulta que mostri tots els usuaris amb més de 80 transaccions utilitzant almenys 2 taules.*/

SELECT *
FROM all_users
WHERE EXISTS (
SELECT user_id, COUNT(id) AS number_transactions
FROM transactions
WHERE all_users.id = transactions.user_id
GROUP BY user_id
HAVING number_transactions > 80);
```

34:285

Result Grid — Filter Rows: Search — Edit: — Export/Import:

id	name	surname	phone	email	birth_date	country	city	postal_code	address	region
185	Molly	Gilliam	0800 120 8023	donec@outlook.couk	Dec 21, 1993	United Kingdom	London	EC1A 1BB	P.O. Box 202, 5638 Mi Rd.	europe
289	Dxwgi	Hwcru	+98-309-8797	dxwgi.hwcru@example.com	Aug 20, 1976	Germany	Stuttgart	70173	82 Hwcru Street	europe
318	Bnyr	Astuw	+33-120-9644	bnyr.astuw@example.com	May 3, 1974	Italy	Genoa	16100	53 Astuw Street	europe
454	Sfzzoh	Xgvfridxs	+58-495-3945	sfzzoh.xgvfridxs@example.com	Aug 28, 1962	Poland	Gdansk	80-001	52 Xgvfridxs Street	europe
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

all_users 4 — Apply — Revert

Action Output

	Time	Action	Response	Duration / Fetch Time
12	10:33:13	ALTER TABLE transactions ADD CONSTRAINT fk_transactions_credit_cards_id FOREIGN KEY (card_id) REFE...	100000 row(s) affected Records: 100000 Duplicates:...	0.706 sec
13	10:39:55	SELECT * FROM all_users WHERE EXISTS (SELECT user_id, COUNT(id) AS number_transactions FROM tran...	4 row(s) returned	0.074 sec / 0.000010...

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

SELECT cc.iban, ROUND(AVG(amount),2) AS average_amount
FROM transactions t
JOIN companies c ON t.business_id = c.company_id
JOIN credit_cards cc ON t.card_id = cc.id
WHERE c.company_name LIKE '%Donec Ltd%' -- Cambio P2P de nombre de compañía, tenía Donec solo y había más
GROUP BY cc.iban;
```

21:299 — 1 error found

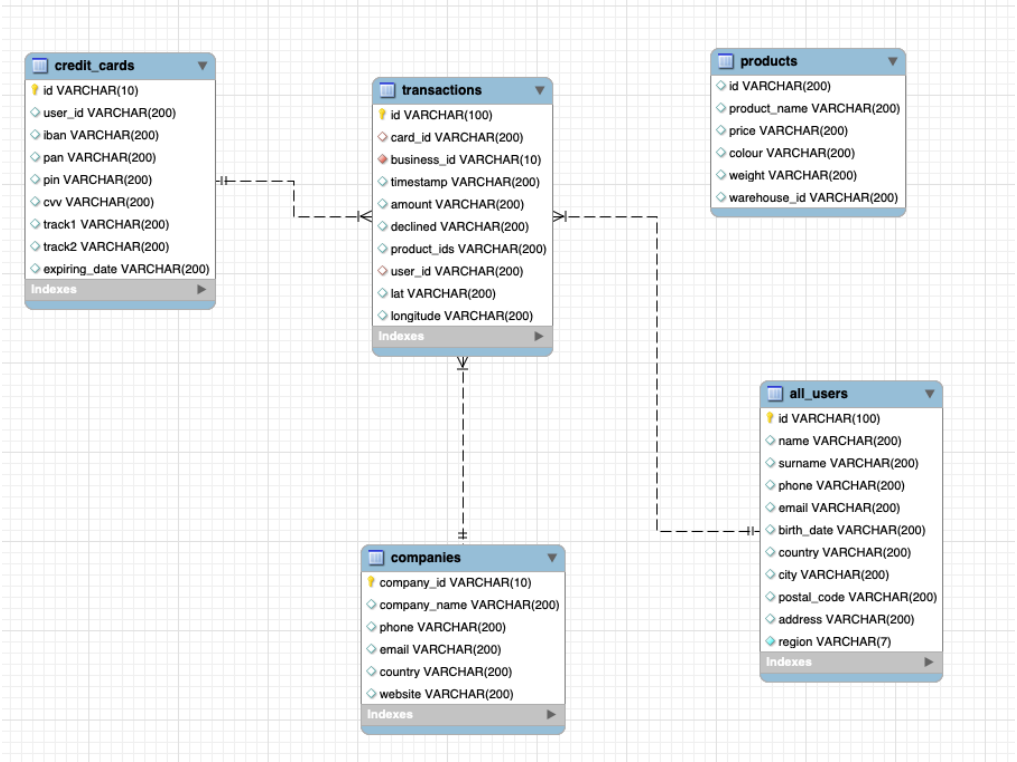
Result Grid — Filter Rows: Search — Export:

iban	average_amount
XX9114064011255863075868 05	356.25
SK9446370242474562577506	142.96
XX7767529178459529755556 40	257.37
XX4138273622897193049089 90	139.59
XX3477872460707696107803 08	240.41

Result 29 — Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
57	13:37:01	ALTER TABLE transactions MODIFY COLUMN id VARCHAR(100) NOT NULL, ADD CONSTRAINT pk_transactions_id...	100000 row(s) affected Records: 100000 Duplicates:...	1.283 sec
58	13:39:55	SELECT cc.iban, ROUND(AVG(amount),2) AS average_amount FROM transactions t JOIN companies c ON t.busines...	371 row(s) returned	0.015 sec / 0.000059...



Nivell 2

Crea una nova taula que reflecteixi l'estat de les targetes de crèdit basat en si les tres últimes transaccions han estat declinades aleshores és inactiu, si almenys una no és rebutjada aleshores és actiu. Partint d'aquesta taula respon:

Exercici 1

Quantes targetes estan actives?

Per començar he mirat la distribució de la taula transaction així com he modificat el tipus de dada de la columna 'timestamp' a timestamp per a poder ordenar correctament.

```
-- Creació taula
SELECT *
FROM transactions;

SELECT card_id, declined, timestamp
FROM transactions
ORDER BY card_id, timestamp;

ALTER TABLE transactions
MODIFY COLUMN timestamp TIMESTAMP;
```

Result Grid

id	card_id	business_id	timestamp	amount	declined	product_ids	user_id	lat	longitude
00043A49-2949-494B-A5DD-A5BAE3BB19DD	CcS-9294	b-2458	2024-08-28 07:16:46	395.43	0	16, 26, 97, 87	4713	46.19992926158272	1.4355402821327607
000447FE-B650-4DCF-85DE-C7ED0EE1CAAD	CcS-5019	b-2370	2016-12-21 20:07:18	155.63	0	66, 69, 87	438	41.59720554463741	12.22175994259365
00045D6B-ED2E-4F2F-8186-CEE074D875D0	CcS-6699	b-2390	2020-07-14 15:37:45	326.01	0	30, 11, 16, 81	2118	29.7572958999964348	-95.37963676492454
000481C3-1C26-4FEF-83A0-4CD0EB004BBD	CcS-6696	b-2230	2017-09-04 19:44:53	161.6	0	72	2115	53.54888376797025	-113.50305274646564
00051AA4-9CBE-4268-B070-C38062A1B3E2	CcS-7606	b-2266	2017-01-05 18:19:25	148.91	0	18	3025	52.20836951654172	5.690806474241335
0008A312-EDFE-4A4F-BC99-E9C92EC3CA4D	CcU-3358	b-2598	2023-09-23 04:51:43	294.59	0	35, 33, 19	215	53.553485560256014	-113.49913398747 81
0009A151-9BCF-4E31-9053-A468FF77FAAB	CcS-7509	b-2546	2023-12-31 00:06:36	383.63	0	93, 55, 28, 91	2928	51.93615735576977	5.342650184655181

transactions 13 | transactions 14

Action Output

	Time	Action	Response	Duration / Fe
11...	13:33:27	SELECT * FROM transactions LIMIT 0, 50000	50000 row(s) returned	0.0011 sec /
11...	13:33:27	SELECT card_id, declined, timestamp FROM transactions ORDER BY card_id, timestamp LIMIT 0, 50000	50000 row(s) returned	0.073 sec / 0
11...	13:33:27	ALTER TABLE transactions MODIFY COLUMN timestamp TIMESTAMP	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0	0.0023 sec

Per tal de crear la taula 'active_credit_cards' he realitzat dues CTE i una consulta final. A la primera CTE he utilitzat la *window function* ROW_NUMBER per tal de saber les últimes tres transaccions per a cada targeta i ordenar-les per temporalitat de manera descendent, a la segona he creat una altra CTE on només apareixien les tres últimes transaccions i a la consulta final he utilitzat un CASE WHEN per a definir els paràmetres 'inactive' i 'active' a través de la suma de declined..

```
CREATE TABLE active_credit_cards AS
    WITH ranked_transactions AS
        (SELECT card_id, declined, transaction_date,
            ROW_NUMBER() OVER(PARTITION BY card_id ORDER BY transaction_date DESC) AS rn
        FROM transactions
        ),
        last3_transactions AS (
            SELECT *
            FROM ranked_transactions
            WHERE rn<= 3
            )
    SELECT card_id,
        CASE WHEN SUM(declined) = 3 THEN 'inactive'
        ELSE 'active'
        END AS state
    FROM last3_transactions
    GROUP BY card_id;
```

Result Grid

active_cards
4995

active_credit_cards 1 | Result 2 | Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
2	07:11:28	CREATE TABLE active_credit_cards AS WITH ranked_transactions AS (SELECT card_id, declined, transacti...	5000 row(s) affected Records: 5000 Duplicates: 0...	0.248 sec
3	07:11:28	SELECT * FROM active_credit_cards LIMIT 0, 50000	5000 row(s) returned	0.00047 sec / 0.0021...
4	07:11:28	SELECT COUNT(card_id) AS active_cards FROM active_credit_cards WHERE state = 'active' LIMIT 0, 50000	1 row(s) returned	0.0016 sec / 0.00000...

Per tal de saber quantes targetes estan actives he fet la següent consulta:

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:

Exercici 1

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

Primer de tot he creat la taula 'transaction_product' utilitzant la funció JSON_TABLE (on primer he necessitat transformar les dades en format array). Posteriorment he comprovat que s'havia creat correctament (visualitzant també quants productes hi havia per cada transacció a la taula transactions (columna product_ids, no a la captura)).

```
CREATE TABLE transaction_product AS
    SELECT
      t.id AS transaction_id,
      jt.product_id
    FROM transactions t
    CROSS JOIN JSON_TABLE(
      CONCAT('[',t.product_ids,']'), -- changed from REPLACE in P2P
      '$[*]' COLUMNS (
        product_id INT PATH '$'
      )
    ) jt;


/*
```

100% 65:425

Action Output

	Time	Action	Response	Duration / Fetch Time
73	13:47:35	CREATE TABLE transaction_product AS SELECT t.id AS transaction_id, jt.product_id FROM transactions t CRO...	253391 row(s) affected Records: 253391 Duplicates:...	0.598 sec

```
SELECT *
FROM transaction_product;
```

100% 12:452

Result Grid Filter Rows: Search Export: Fetch rows:

transaction_id	product_id
00043A49-2949-494B-A5DD-A5BAE3BB19DD	16
00043A49-2949-494B-A5DD-A5BAE3BB19DD	26
00043A49-2949-494B-A5DD-A5BAE3BB19DD	97
00043A49-2949-494B-A5DD-A5BAE3BB19DD	87
000447FE-B650-4DCF-85DE-C7ED0EE1CAAD	66

transaction_product 36 Read Only

Action Output

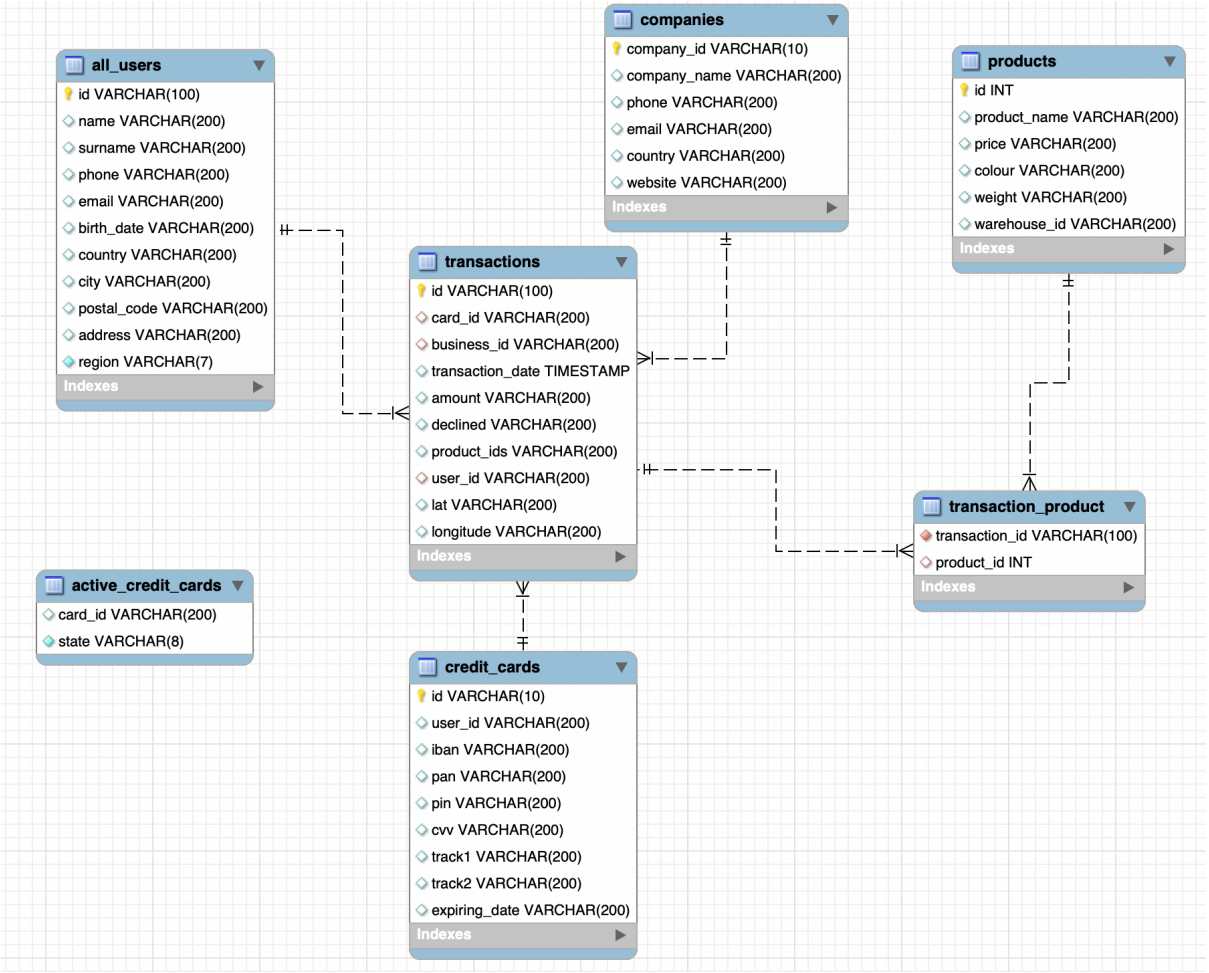
	Time	Action	Response	Duration / Fetch Time
74	13:48:45	SELECT * FROM transaction_product	253391 row(s) returned	0.00100 sec / 0.106 s...

Posteriorment, per tal d'associar la taula intermèdia amb les altres taules he definit a la taula 'products' la columna 'id' com a PK (canviant el tipus de dada abans) mentre que a la taula creada 'transaction_products' he definit com a FK les dues columnes, relacionant-les amb les taules corresponents.

```
ALTER TABLE products
MODIFY COLUMN id INT NOT NULL, -- needs to be the same datatype as intermediate table
ADD CONSTRAINT pk_products_id PRIMARY KEY (id);

-- ALTER TABLE products
-- DROP PRIMARY KEY;


ALTER TABLE transaction_product
ADD CONSTRAINT fk_int_transaction_id
FOREIGN KEY (transaction_id) REFERENCES transactions(id),
ADD CONSTRAINT fk_int_product_id
FOREIGN KEY (product_id) REFERENCES products(id);
```

	Time	Action	Response	Duration / Fetch Time
65	11:40:36	ALTER TABLE products MODIFY COLUMN id INT NOT NU...	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0	0.025 sec
66	11:40:36	ALTER TABLE transaction_product ADD CONSTRAINT fk_...	253391 row(s) affected Records: 253391 Duplicates: 0 Warnings: 0	1.050 sec

El diagrama de relacions final és el següent:



La taula ‘active_credit_cards’ no l’he associat perquè pot variar cada cop que entrin noves dades. Si s’hagués de fer més cops es podria crear una vista per tal de tenir-hi accés de forma senzilla i també poder garantir l’accés a la consulta a determinades persones.

Per tal de saber quantes unitats s’han venut de cada producte, he realitzat la següent query. He comprovar que hi hagi el mateix nombre de productes que a la taula productes:

```
SELECT product_id, COUNT(transaction_id) AS number_sold
FROM transaction_product
GROUP BY product_id
ORDER BY product_id;

SELECT *
FROM products;
```

Result Grid | Filter Rows: Search | Export:

product_id	number_sold
1	2468
2	2573
3	2528
4	2584
5	2548
6	2492

Result 32 | products 33 | Read Only

	Time	Action	Response	Duration / Fetch Time
54	11:25:03	SELECT product_id, COUNT(transaction_id) AS number_s...	100 row(s) returned	0.099 sec / 0.00006...
55	11:25:03	SELECT * FROM products LIMIT 0, 50000	100 row(s) returned	0.00054 sec / 0.000...