SPRINT 4: Creació de Base de Dades

Elaborat per:

Vanesa Pérez Ramírez

Data: 08/10/2025

SPRINT 4	1: Creació de Base de Dades	1
Nivell	1	3
•	Elaboració prèvia del model.	3
•	Creació de la base de dades:	4
•	Creació de taules	4
•	Càrrega de les dades a totes les taules:	5
•	Modificació de dades de les taules:	6
•	Taula users:	7
•	Taula companies:	10
•	Taula crèdit_cards:	12
•	Taula transactions:	14
•	Creació de les PK a cada taula:	15
•	Creació de les PK a cada taula:	16
Exerci	ici 1	17
Exerci	ici 2	18
Nivell	2	19
Exerci	ici 1	21
Nivell	3	22
•	Càrrega i modelatge de la taula products	22
•	Creació de la taula intermitja: products.transactions	23
Exerci	ici 1	26

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:

Per crear la base de dades des de arxius .csv seguiré les indicacion de la article que ens han proporcionat a l'sprint 4 del present curs:

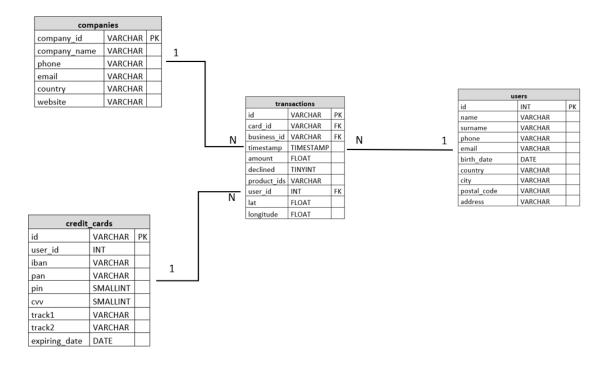
https://www.delftstack.com/es/howto/mysql/mysql-import-csv/



Primer he revisat cada un dels arxius i per la informació que obtinc el disseny previ de la base de dades que he fet és, el lenght dels tipus de dades el determinaré en funció de les dades que obtindré un cop carregat els arxius. Per altra banda, ajuntaré les taules american_users i european_users carregant les dades de cada taula a una nova taula que anomenaré users, per tenir tots els users a la mateixa taula, sí no es fes així la fk amb la taula transactions no es vincularia amb les taules del users de cada continent, donaria error perque haurien id a transactions que no existirian a les taules americans o europeans.

Elaboració prèvia del model.

El model previst es visualitza de la següent forma:



• Creació de la base de dades:

Es crea la base de dades nova, el seu nom serà ventas i es selcciona com a BBDD amb la que anem a treballar:

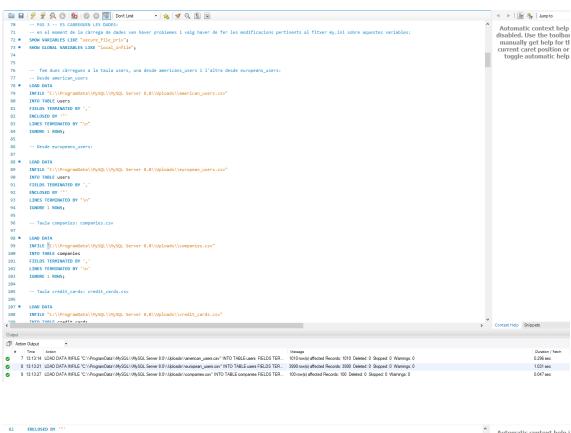
```
Affinition to content help in a solution of the content help in a solution
```

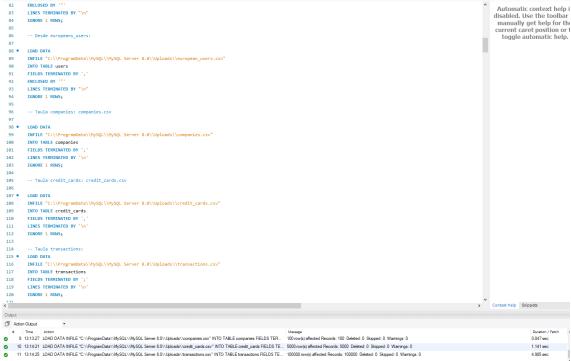
Creació de taules

Es creen les taules: users, companyies, crèdit_cards i transactions. Al seguir les indicacions de l'article llegit, per carregar les dades es fa toto en VARCHAR (255) i un cop carregades modificarem el tipus de dades i les dades si cal.

```
id VARCHAR(255) NULL,
                       me VARCHAR(255) NULL,
                   surname VARCHAR(255) NULL,
                   phone VARCHAR(255) NULL,
email VARCHAR(255) NULL,
                 birth_date VARCHAR(255) NULL,
                  country VARCHAR(255) NULL,
                  city VARCHAR(255) NULL,
postal_code VARCHAR(255) NULL,
                   address VARCHAR(255) NULL
      34 • ⊖ CREATE TABLE IF NOT EXISTS companies (
                  company_id VARCHAR(255) NULL,
                  company_name VARCHAR(255) NULL,
phone VARCHAR(255) NULL,
email VARCHAR(255) NULL,
                  country VARCHAR(255) NULL,
website VARCHAR(255) NULL
      41
      43 \bullet \ominus CREATE TABLE IF NOT EXISTS credit_cards (
                  id VARCHAR(255) NULL,
                   user_id VARCHAR(255) NULL,
                  iban VARCHAR(255) NULL,
pan VARCHAR(255) NULL,
                 pin VARCHAR(255) NULL,
cvv VARCHAR(255) NULL,
                  track1 VARCHAR(255) NULL,
track2 VARCHAR(255) NULL,
                   expiring_date VARCHAR(255) NULL
                                                                                                                                                                                                         toggle automatic help.
            ○ CREATE TABLE IF NOT EXISTS companies (
company_id VARCHAR(255) NULL,
company_name VARCHAR(255) NULL,
phone VARCHAR(255) NULL,
email VARCHAR(255) NULL,
country VARCHAR(255) NULL,
website VARCHAR(255) NULL
```

• Càrrega de les dades a totes les taules:





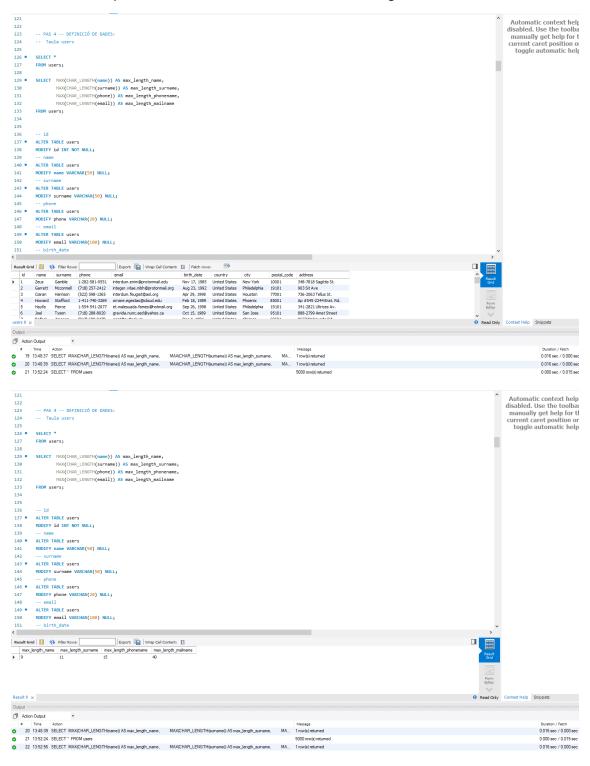
Modificació de dades de les taules:

Un cop carregades les dades comencem a modificar-les per posar el tipus de dades corresponents. El que faré és mirar el tipus de dada i el max(lenght()) de les dades string i a través de l'UPDATE modificaré cada columna .

Com si ho mostro dada per dada no és aclaratori, el que faré és posar dos exemples de modificació de dades i a les columnes que s'hagi de tenir en compte algun detall important o posaré

• Taula users:

Es fa un select * per veure totes les dades i es busca el max lenght de les dades:



Les decisions preses de modificació:

Id: INT acabarà sent la Primary Key per tant NOT NULL

<u>Name</u>: mirem la max longitud de l'string, i decideixo posar 50. Es un string i poso VARCHAR (50)

Surname: aplico la mateixa lògica que a name.

<u>Phone:</u> al mirar les dades phone veiem que hi ha una barreja de números i caràcters per tant les dades seran VARCHAR. El lenght màxim es 14, posarem un VARCHAR (20)

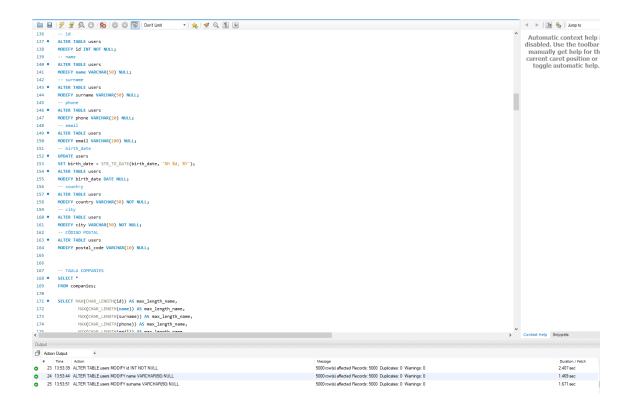
<u>Mail:</u> podem veure que es un string i que la longitud màxima es de 40 caracters, però decideixo canvia a un VARCHAR (100) com a tipus de valor, ja que el emails poden ser encara mes llargs.

<u>Birth_date:</u> quan observem les dates veiem que es un format string. He fet la proba de passarho a DATE i no deixa no detecta el format, per tant, hem de modificar la columna birth_date per tal d'arribar a tenir les dates de naixement en un format DATE. El que he fet és modificar la columna per a que el format de string sigui un fomat de data y he substituït (UPDATE).

Country: ho modifiquem seguint el procés que hem dut a terme amb el camp mail:

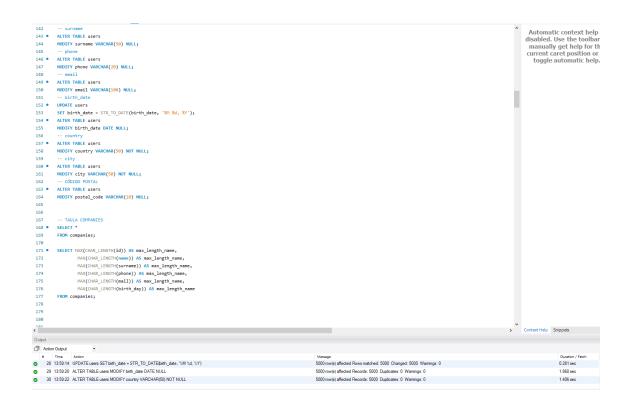
<u>Postal_code</u>: ho modifiquem seguint el procés que hem dut a terme amb el camp mail:

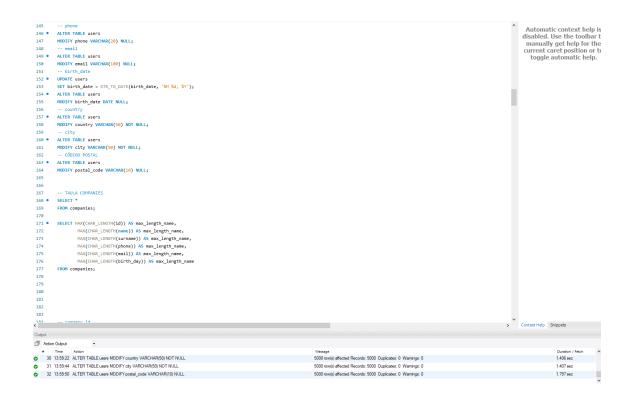
Adress: ho deizarem com està ja que les adresses són molt diverses.



```
Automatic context help disabled, tise the toolbar manually set help for the current caret position or toggle automatic help.

A THE TABLE SET THE SET
```

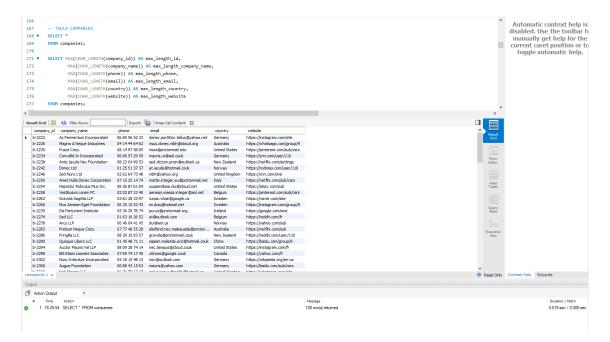




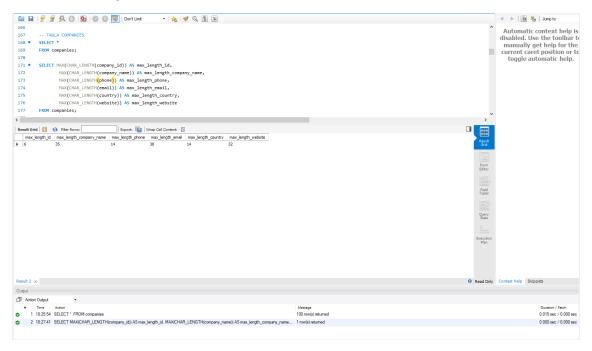
Taula companies:

Seguirem el mateix procés que amb les altres taules:

Estudiar les dades:



Veure el Max(length()) de les dades:



Company id

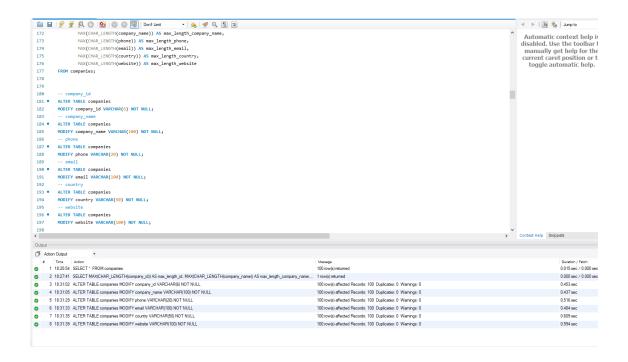
Company name

<u>Phone:</u> He decidit posar el mateix tipus de dada que el telèfon del users per unificar criteris. Encara que en aquesta taula són tots de la mateixa longitud i la mateixa estructura, podíem haver posar INT però he preferid unificar criteris així que he posat VARCHAR(20):

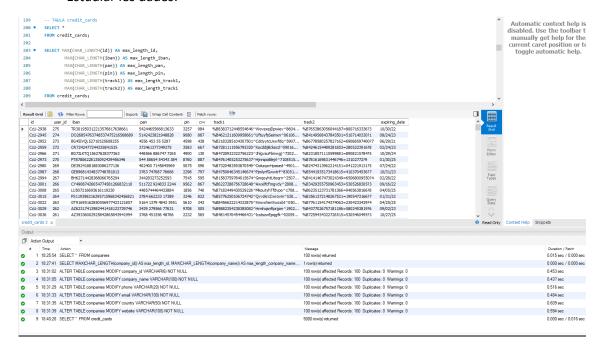
Email: he unificat criteris i he posat el mateix tipus de dades que a la taula users

Country: he unificat criteris i he posat el mateix tipus de dades que a la taula users

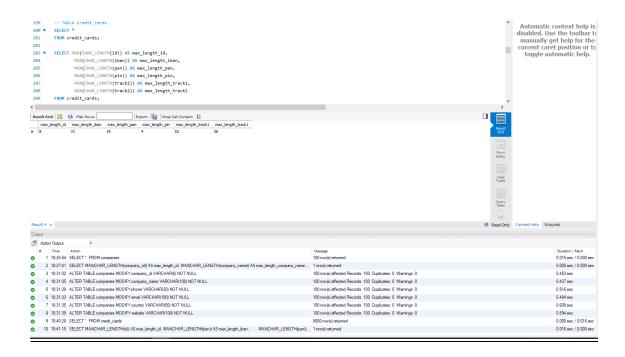
Website



- Taula crèdit_cards:
- Seguirem el mateix procés que amb les altres taules:
- Estudiar les dades:



Veure el Max(length()) de les dades:



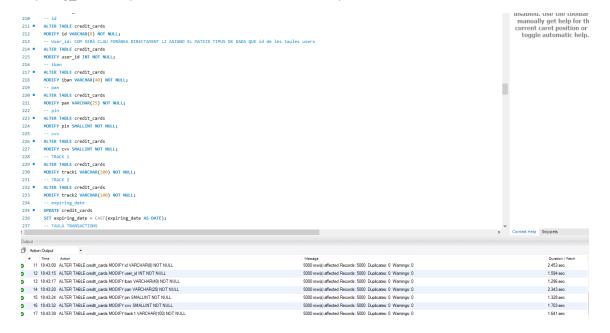
<u>Id</u>

<u>User id:</u> com serà FK de les taultes users poso directament el tipus de dades d'aquell camp.

IBAN: He seguit el procés.

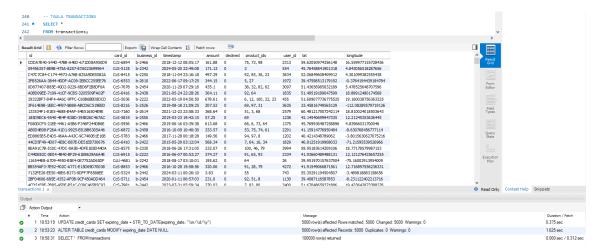
Pan: era sèrie de números amb espais he posat VARCHAR. Havía fet la proba de treure els espais i posar un BIGINT però finalment he decidit que si s'actualitzan les dades amb bases que vinguin de fora i limito a BIGINT potser que donés problemes.

Expiring date: he pasat el valor de STR TO DATE i després ho he modificat a date.

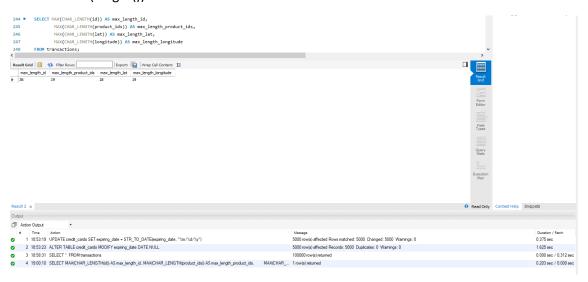




- <u>Taula transactions:</u>
- Seguirem el mateix procés que amb les altres taules:
- Estudiar les dades:



Veure el Max(length()) de les dades:



Id

Card_id: al ser FK el tipus de dades serà el mateix que id a credit_cards

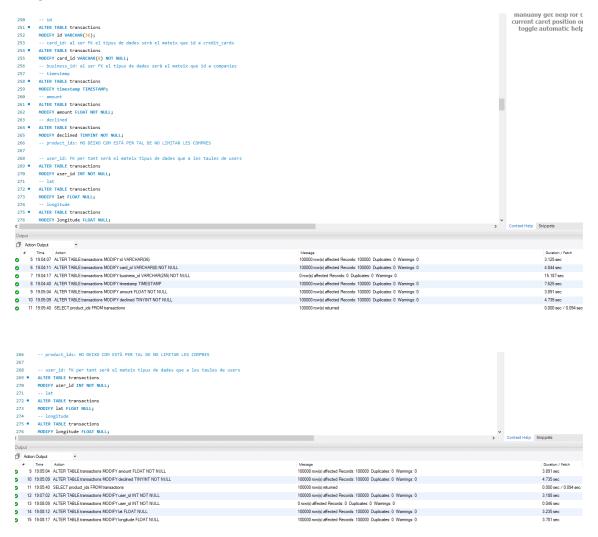
Business_id: al ser FK el tipus de dades serà el mateix que id a companies

Product_ids: ho deixo com està per tal de no limitar les compres

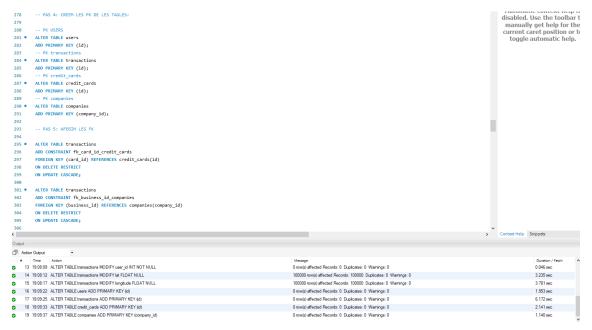
User_id: FK per tant serà el mateix tipus de dades que a les taules de users

Lat

Longitude



• Creació de les PK a cada taula:

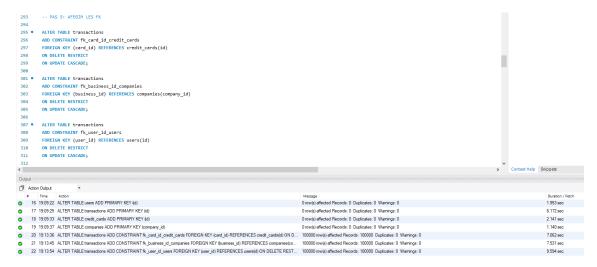


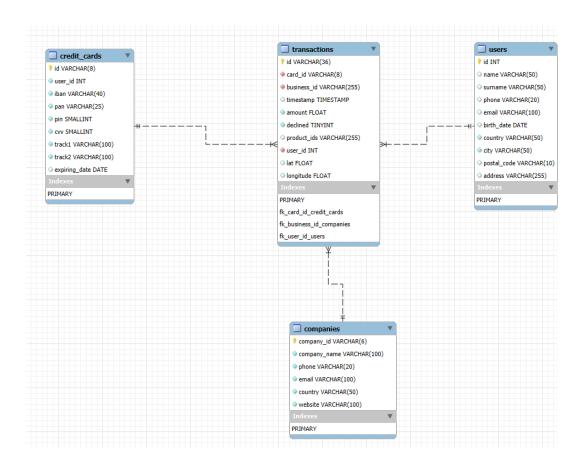
• Creació de les PK a cada taula:

Afegim les Foreing Key a la taula de fets transactions. Les Foreing Key son:

- card_id amb la taula crèdit_cards
- business_id amb la taula company
- user_id amb la taula users

Durant la creació de les PK decideixo que en el cas de que es faci un UPDATE a la PK de les taules mare aquesta actualització es traslladi a les taules filles i que en cas de que es vulguin liminar registres de PK no s'eliminin.





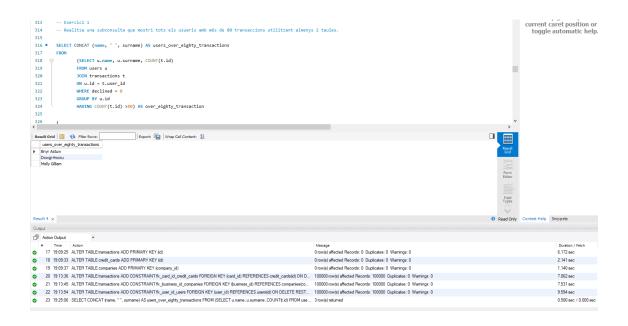
Finalment, el diagrama que ha quedat ha sigut el següent:

Observem que les taules companies, crèdit_cards i users estan relacionades a través de la seva PK amb la taula transactions i que les relacions que se estableixen són de 1:N

Exercici 1

Realitza una subconsulta que mostri tots els usuaris amb més de 80 transaccions utilitzant almenys 2 taules.

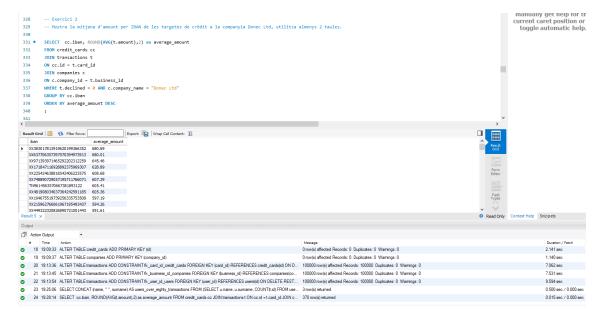
A la subconsulta s'utilitzen dues taules i mostra els usuaris amb més transaccions. Si la he de transformar en subconsulta que és el que demana l'enunciat de l'exercici, el que faré es crear una consulta que concateni nom i cognom i en el FROM posaré la subconsulta:



Exercici 2

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

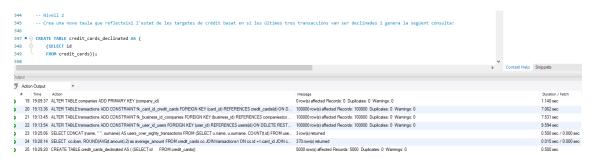
Faig un GROUP BY per iban amb la funció d'agregació AVG() i en utilitzo tres taules fent JOIN, he decidit ordenar pels valors de mitjana de major a menor:



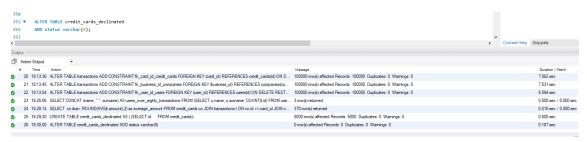
Nivell 2

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:

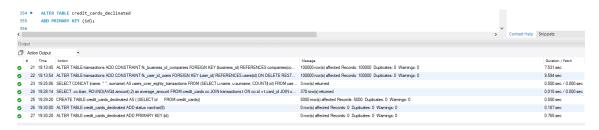
He creat la nova taula desde credit card selección les id de les credit cards.



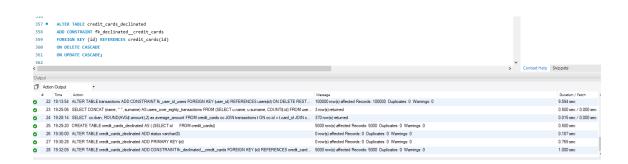
He afegit la columna status a on s'indicaran si las targetes están actives o inactives



He creat la PK

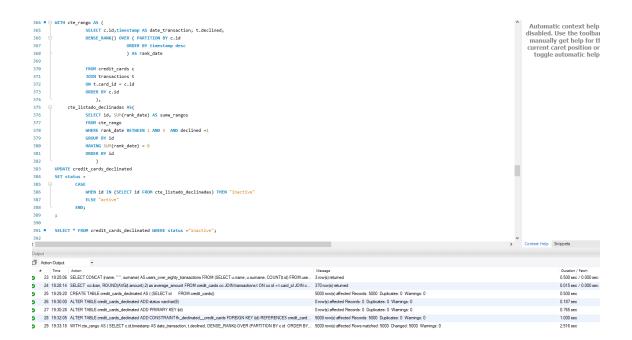


He creat la FK amb la taula credit cards:

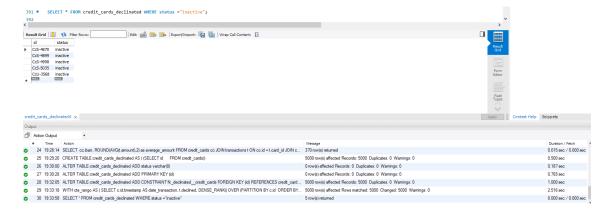


Per afegir les dades a la columnes he creat una cosulta Update amb cte's:

- cte_rango: a on poso rank a les dates de transaccions per id, ho he fet amb una finestra de dense_rank i amb PARTITION BY, l'ordre de resultat es per id i per rank_date.
- cte_listado_declinadas: una consulta que em dona el llista de les targetes declinades en les tres darreres transaccions, que són a les que els hi posaré l'estatus de inactiva.
 En aquesta CTE per poder filtrar les id's he filtrat per rank i per declined = 0 y he fet un group by de les id's en les que es sumen el ranks que apareixen del filtre, i he posat un filtre having al group by en el que filtro que si la suma dels valors del rank son 6. La lògica que he aplicat es que si una targeta ha sigut declinada les tres darreres dates s'ha desactivat per tant la suma de tres ranks desactivats son 6.
 En aquest cas, he utilitzat la cte_rango per fer els filtres
- Per finalitzar, faré la consulta d' UPDATE en la que li diré que en el cas en el que l'id estigui inclòs al llistat de id en el que les targetes estan fora de servei (cte_listado_declinadas) posi inactive i en el cas contrari posi active.



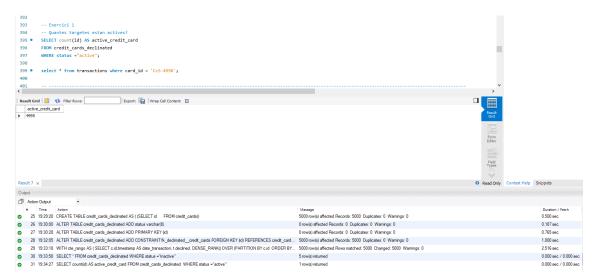
Ho comprovo:



Exercici 1

Quantes targetes estan actives?

He aplicat una consulta en la que es conten el número de targes on la condició és que han d'estar actives:



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:

Càrrega i modelatge de la taula products

Seguiré els mateixos passos que s'ha fet per crear el model d'estrella:

Creo i carrego la taula:



Veiem com són les dades:



Veiem longitud de les dades

I fem les següents modificacions per tal de tenir els tipus de dades desitjats:

Id: el posarem INT NOT NULL

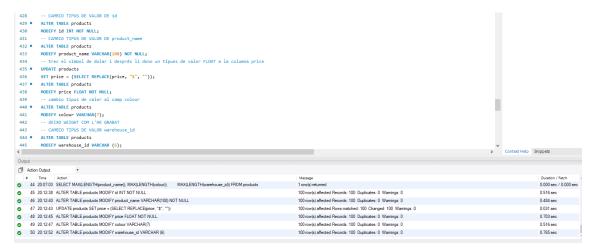
Product_name: VARCHAR més curt

Price: trèiem el símbol \$ i el pasem a FLOAT

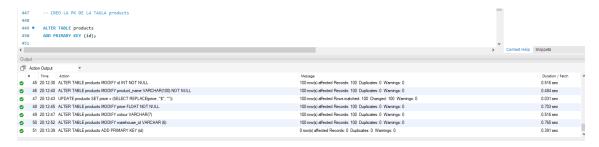
Colour: VARCHAR més curt

Deixo weight com l'he grabat

warehouse_id: VARCHAR més curt



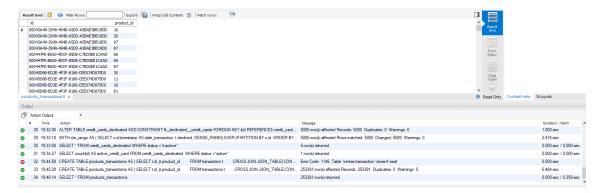
Creo la PK:



Creació de la taula intermitja: products.transactions

Inicialment, la taula la he creat utilitzant la funció Length() i la funció SUBSTRING_INDEX(). Donava correctament, però durant la correcció de l'exercici m'han explicat que la miller manera de realitzar-lo és utilitzant la funció JSON_TABLE(), i finalment he decidit fer-ho amb aquesta funció. Per completar la informació de la decisió he de dir que als apunts indiquen que per fer taules Intermitges és millor utilitzar aquesta funció.

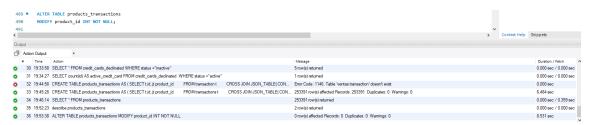
Veiem com queda la taula



Comprovem com han quedat el tipus de dades:



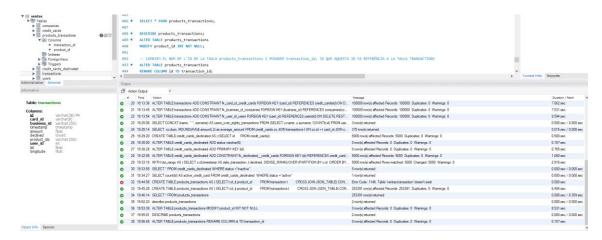
Decideixo posar que product_id no tingui Null:



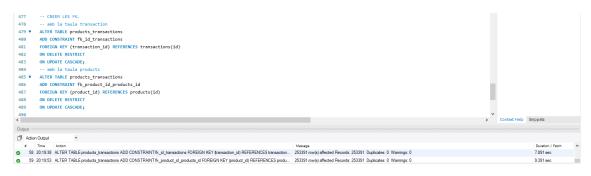
Tornem a executar el DESCRIBE



Cambiem el nom de l'id pel de transaction_id, per que quedi clar a que es refereix aquest camp:

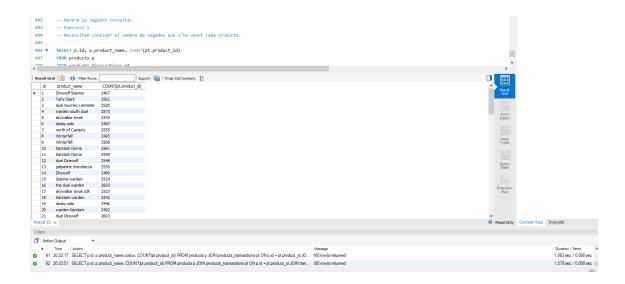


Creem les Foreing key: amb la taula transaction i amb la taula products



Exercici 1

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



Durant la revisió de l'exercici hem vist que donva correcte però que hi havia noms de productes repetits i hem fet la proba afegint les altres característiques que defineixen el producte i veiem que encara que el nom és el mateix l'id del producte diferència no només el nom si no les altres característiques:

